

Mandy Lange

Partielle Korrelationen und Partial Mutual Information zur Analyse von fMRT-Zeitreihen

MASTERARBEIT

Hochschule Mittweida

University of Applied Sciences

Fachbereich Mathematik/Naturwissenschaften/Informatik

Mittweida, November 2011

Mandy Lange

Partielle Korrelationen und Partial Mutual Information zur Analyse von fMRT-Zeitreihen

eingereicht als

MASTERARBEIT

an der

Hochschule Mittweida

University of Applied Sciences

Fachbereich Mathematik/Naturwissenschaften/Informatik

Mittweida, 2011

Erstprüfer: Prof. Dr. habil. Villmann

Zweitprüfer: PD. Dr. Lohmann

Vorgelegte Arbeit wurde verteidigt am: 30.11.2011

Inhaltsverzeichnis

1	Zielsetzung	2
2	Der fMRT-Datensatz	5
2.1	Die Entstehung eines fMRT-Datensatzes	5
2.2	Charakterisierung des hier verwendeten Datensatzes	6
3	Eigenvektor Zentralität	8
3.1	Zentralitäten in Graphen	9
3.2	Berechnung der Eigenvektor Zentralität	10
4	Vektorquantisierung	13
4.1	Einleitung	13
4.2	Prinzip	14
4.3	Der Neural Gas Algorithmus	14
5	Clustering der fMRI Prototypen	19
5.1	Einleitung	19
5.2	Anwendung auf die fMRT-Prototypen	20
6	Partielle Korrelationen	24
6.1	Der Korrelationskoeffizient	24
6.1.1	Korrelation nach Pearson	25
6.1.2	Rangkorrelationskoeffizient nach Spearman	27
6.1.3	Rangkorrelationskoeffizient nach Kendall	28
6.1.4	Signifikanztests	29
6.2	Bedingte und Partielle Korrelationen	30
6.3	Partielle Korrelationen höherer Ordnung	34
6.3.1	Berechnung der partiellen Korrelation mit linearer Regression	35

6.3.2	Rekursive Berechnung der partielle Korrelation	35
6.3.3	Verwendung der inversen Korrelationsmatrix	36
6.3.4	Interpretation von Korrelationen	38
6.3.5	Ein Scheinkorrelationsmaß	40
7	Partial Mutual Information	42
7.1	Informationstheoretische Begriffe	42
7.2	Mutual Information	47
7.3	Partial Mutual Information	48
7.3.1	Schätzung der Partial Mutual Information	50
7.3.1.1	Entropieschätzung	50
7.3.1.2	Entropieschätzung basierend auf den k -nächsten Nachbar	53
7.3.2	Eine Ersatzmethode für den Signifikanztest	56
7.3.3	Pseudo Mutual Information	57
7.3.4	Skalierungsmethode	58
7.4	Partial Mutual Information höherer Ordnung	58
7.4.1	Die Berechnung der Partial Mutual Information höherer Ordnung	61
8	Auswertung und Ergebnisse	63
8.1	Vergleich der partiellen Korrelationsmatrix und der Partial Mutual Infor- mation Matrix	63
8.2	Auswertung der Eigenvektor Zentralität	64
9	Zusammenfassung und Diskussion	71
A	Notationen	75
	Abbildungsverzeichnis	77
	Literaturverzeichnis	80

Selbstständigkeitserklärung

Ich versichere wahrheitsgemäß, die Masterarbeit selbständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

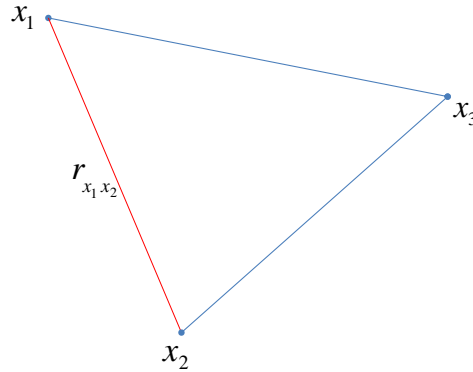
Mittweida, den 23.11.2011

1 Zielsetzung

Für die wissenschaftliche Untersuchung der kognitiven Fähigkeiten und Gehirnprozesse beim Menschen können mit dem Magnet-Resonanz-Tomograph (MRT) funktionelle Daten vom Gehirn aufgenommen werden. Dabei entsteht für jeden Messpunkt eine Zeitreihe, das fMRT-Signal (das funktionelle Magnet-Resonanz-Tomograph-Signal). Dieser Messpunkt wird auch als Voxel bezeichnet. Ein Voxel ist ein dreidimensionaler Datenpunkt, das heißt, es existiert zu jedem Voxel die Koordinate (x, y, z) . Eine Menge von Voxels, die untereinander verbunden sind, bilden ein Voxelgitter, welches bei der Datenrepräsentation vor allem in den bildgebenden Verfahren, wie der Magnet Resonanz Tomographie, zum Einsatz kommt. Die Länge der Zeitreihe in jedem Voxel entspricht der gesamten Aufnahmedauer.[24]

Von besonderem Interesse ist der funktionelle Zusammenhang der Gehirnregionen. Dabei wird das Gehirn graphentheoretisch betrachtet, dass heißt, das Voxelgitter wird durch einen ungerichteten Graphen modelliert. Die Voxel sind die Knoten in einem ungerichteten Graphen. Die Verbindung zwischen je zwei Knoten sind die Kanten in dem Graphen und das Kantengewicht ist der gewöhnliche Korrelationskoeffizient zwischen jeweils zwei Knoten. In dem so resultierenden Netzwerk vom Gehirn sind direkte und indirekte Verbindungen enthalten. Wenn zwei Voxel von weiteren Voxeln beeinflusst werden und aufgrund dessen eine hoher Korrelationskoeffizient entsteht, dann nennt man die Verbindung zwischen diesen zwei Voxeln indirekt, siehe Abbildung 1.0.1. Um die Netzwerkstruktur vom Gehirn sichtbar zu machen, gilt es die indirekten Verbindungen zu beseitigen.

In der Arbeit von G. Marralec [21] wurden die marginale (gewöhnliche) Korrelation und die partielle Korrelation, welche im Kapitel 6 erklärt werden, zur Erkennung von Netzwerkstrukturen gegenübergestellt. In dieser Studie wurden sechs spezielle Regionen vom Gehirn ausgewählt und davon die Korrelationsmatrix und die partielle Korrelations-



1

Figure 1.0.1: Diese Abbildung stellt eine indirekte schematische Verbindung dar:
 Ein hoher Korrelationskoeffizient $r_{x_1 x_2}$ zwischen den Voxeln x_1 und x_2 kann aufgrund des Einflusses von Voxel x_3 entstehen.

matrix bestimmt. Diese Matrizen sollen ein Indikator für die Zusammenhangsstruktur der sechs Gehirnregionen sein. Dabei wurde untersucht, ob sich in den Matrizen ein signifikanter Unterschied ergibt, wenn der Proband eine linke oder rechte Handbewegung ausführt. Allerdings wurde in dieser Studie nicht darauf eingegangen, welche Ordnung für die partielle Korrelation zu wählen ist. Man hat die partielle Korrelation zwischen jeweils zwei Gehirnregionen unter Ausschluss der restlichen Mengen berechnet. Dass heißt, es wurde zwischen jeweils zwei Gehirnregionen der Einfluss der restlichen Gehirnregionen herauspartialisiert ohne zu beurteilen, ob der daraus resultierende partielle Korrelationskoeffizient der Ordnung vier geeignet ist. Denn es besteht die Möglichkeit, dass zwei Gehirnregionen von weniger als vier Gehirnregionen beeinflusst werden. Das Ziel der vorliegenden Arbeit ist es also, Netzwerkstrukturen im Gehirn auf der Basis der Korrelationen zu erkennen, in dem ein Maß entwickelt werden soll, welches aussagt, wann eine Scheinkorrelation und damit eine indirekte Verbindung vorliegt. Dabei werden die partiellen Korrelationen zwischen jeweils zwei Variablen (fMRT-Zeitreihen) und jeder möglichen Kombination der restlichen Menge bestimmt.

Aufgrund des außerordentlichen Berechnungsaufwandes für das gesamte Gehirn konnte

die partielle Korrelation bisher nur auf kleine Datensätze, somit nur auf Teile von Gehirnnarealen angewendet werden. [20].

In dieser Arbeit soll die Bestimmung der partiellen Korrelation auf der vollständigen Datenmenge erfolgen. Dafür werden die Daten durch einen kleineren Datensatz repräsentiert. Diese Komprimierung soll es ermöglichen die partiellen Korrelationen auf dem gesamten Datensatz zu bestimmen.

Ein weiteres Ziel ist die Identifizierung von speziellen Knoten, die eine “zentrale” Rolle im Gehirn spielen. Dafür gibt es bestimmte sogenannte Zentralitätsmaße, speziell die Eigenvektor Zentralität. Die Basis solcher Maße ist eine symmetrische Matrix, die für jedes Knotenpaar die Ähnlichkeit angibt. Bisher wurde dafür die oben erwähnte Korrelationsmatrix verwendet, welche auf der Grundlage des gesamten Datensatzes sehr hohe Korrelationen enthält. [19] Damit könnte eine mögliche Verfälschung der Eigenvektor Zentralität entstehen. Es soll nun das Zentralitätsmaß auf der Basis einer neuen Korrelationsmatrix bestimmt werden, die keine Scheinkorrelationen beinhaltet. Damit könnte es eventuell eine klarere Abgrenzung der “zentralen” Knoten im Gehirn geben.

Weiter ist unklar, welche Form der Abhängigkeit zwischen den fMRT-Zeitreihen vorliegt. Da die Korrelation nur ein Maß für lineare Abhängigkeiten ist, wird außerdem die Partial Mutual Information für nicht lineare Abhängigkeiten analysiert und auch als Basis für die Bestimmung von Zentralitäten verwendet.

2 Der fMRT-Datensatz

In diesem Kapitel wird erst kurz beschrieben, wie fMRT-Datensätze generiert werden und es werden die entsprechenden notwendigen Vorverarbeitungsschritte erläutert. Anschließend wird der hier verwendete Datensatz charakterisiert.

2.1 Die Entstehung eines fMRT-Datensatzes

Dieser Abschnitt bezieht sich ausschließlich auf die Literatur [24].

Die Entstehung der Messwerte eines fMRT-Experimentes erfolgt durch die Aufnahme einer Reihe von Bildern mit dem MRT vom Gehirn, welche das BOLD-Signal (blood oxygenation level-dependent signal) repräsentieren. Das Signal beruht auf den unterschiedlichen magnetischen Eigenschaften des sauerstoffreichen beziehungsweise sauerstoffarmen Blutes. Eine gesteigerte Aktivität von Nervenzellen führt letztlich zu einem erhöhten BOLD-Signal in den entsprechenden Hirnregionen.

Moderne Verfahren des fMRT ermöglichen es, in dichter Folge eine große Anzahl von Bildern aufzunehmen, welche das BOLD-Signal darstellen. Dabei ist die Zeit zwischen zwei Aufnahmen konstant und liegt meistens bei 2 Sekunden. Da mittlere Experimentendauer 30 Minuten beträgt, resultieren damit 900 Bilder. Betrachtet man nur ein einziges Voxel, erhält man durch die wiederholte Messung eine BOLD-Zeitreihe, die fMRT-Zeitreihe, die den Wert des Voxels zu jedem Zeitpunkt des Experimentes darstellt, siehe Abbildung 2.1.1. Die Zahl der betrachteten Messpunkte ist hierbei durch die Auflösung der gewonnenen Aufnahmen bestimmt. Bei einer üblichen räumlichen Auflösung von 64×64 Bildpunkten pro Schicht und 12 Schichten würde man knapp 50000 Messpunkte (Voxel) betrachten. Dabei sind Voxelkoordinaten die sogenannten Talairach- oder MNI-Koordinaten. Mit den sogenannten Talairach-Koordinaten lässt sich die dreidimensionale Struktur des Gehirns durch ein Koordinatensystem mit drei Achsen beschreiben. Diese Talairach-Koordinaten geben eine Richtlinie zur Lage der Achsen und zur Standardisierung der Größe

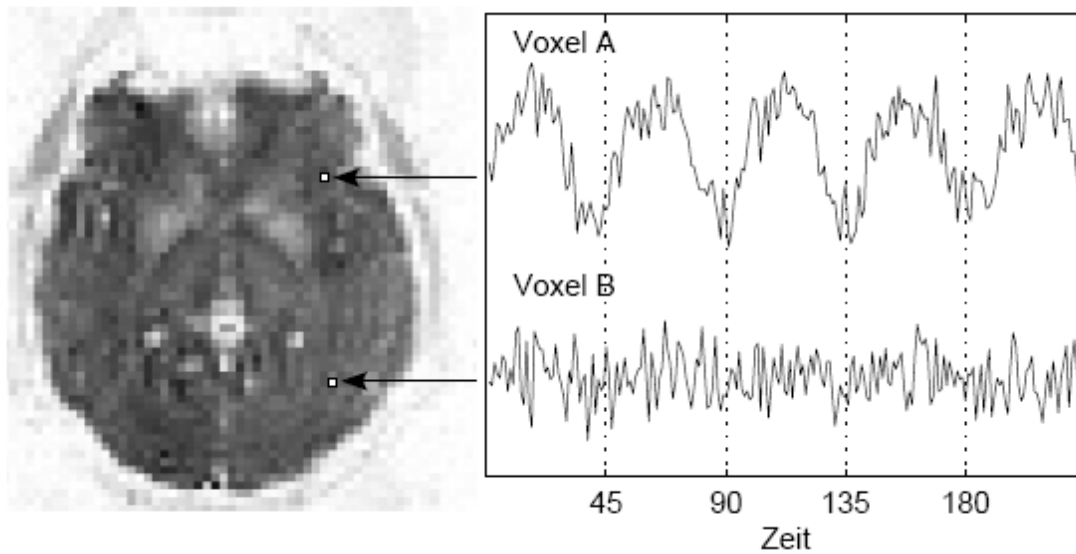


Abb. 2.1.1: BOLD-Zeitreihen von einer aktivierten bzw. nicht aktivierten Hirnregion:
Die Zeitreihe im Voxel *A* entspricht einer aktivierten Hirnregion und die Zeitreihe im Voxel *B* einer nicht aktivierten Hirnregion.

individuell unterschiedlicher Gehirne. Die MNI-Koordinaten können sich mit speziellen Umrechnungen aus den Talairach-Koordinaten ergeben.[7] Die Abkürzung MNI steht für Montral Neurological Institut und begründet sich darauf, dass dieses Institut dieses spezielle Koordinatensystem für die Hirnforschung entwickelt hat.

Vor der Analyse von fMRT-Daten erfolgt in der Regel eine Vorverarbeitung der Daten, um die Datenqualität zu erhöhen und verschiedene technisch oder physiologisch bedingte Artefakte zu beseitigen. Solche Artefakte, die aus dem Signal herausgefiltert werden sollen sind zum Beispiel Herzschlag, Atmung des Probanden oder Drifts, die im Verlauf eines Experimentes entstehen (siehe [18]).

2.2 Charakterisierung des hier verwendeten Datensatzes

Alle Ergebnisse und Untersuchungen in der vorliegenden Arbeit beziehen sich auf den gleichen Datensatz. Wie oben dargelegt, wird hier ein fMRT-Datensatz mit 39460 Zeitreihen (Voxeln) der Länge 168 verwendet. Die Aufnahmedauer war ungefähr fünf Minuten.

Dieser Datensatz stammt aus einer Studie mit 22 Probanden. Im Laufe des Experimentes, währenddessen der verwendete Datensatz entstand, wurden dem Probanden keine Stimuli präsentiert. Die Probanden sind lediglich hungrig oder satt zum Experiment erschienen. Durch diese fehlende Reizpräsentation spiegeln die Messwerte den “resting state” Zustand im Gehirn wieder. [19]

3 Eigenvektor Zentralität

Die Theorie der Eigenvektor Zentralität dieses Kapitels basiert im Wesentlichen auf der Veröffentlichung von G. Lohmann [19].

Die Erforschung der Aktivität bestimmter Gehirnareale kann mit der Magnetresonanztomographie erfolgen. Dieses Verfahren misst den Sauerstoffgehalt des Blutes im Gehirn. Durch den Gebrauch dieser Kenngröße können Rückschlüsse auf die Aktivitäten einzelner Hirnregionen gezogen werden. Für die Auswertung solcher Daten aus funktioneller Magnetresonanztomografie (fMRT) benötigt man mathematische Modelle. Die Auswertung von Stimulus-Response-Experimenten ist relativ einfach, denn die Probanden bekommen während des Experimentes Aufgaben gestellt. Unterdessen misst man die Sauerstoffanreicherung im Blut, welches sich im Verlauf des Experimentes ändert. Diese Änderung des Sauerstoffgehaltes wird verwendet, um signifikante Unterschiede in den Messwerten ausfindig zu machen. Diese weisen darauf hin, dass diese Gehirnareale für die Bewältigung der jeweils gewählten Aufgabe verantwortlich sind. Wenn jedoch dem Probanden keine direkte Aufgabe gestellt wird und er die einzige Aufgabe hat, entweder hungrig oder satt zum Experiment zu erscheinen, ist die Auswertung anspruchsvoller. Dann kann man die Messwerte nicht unmittelbar auf einen experimentellen Stimulus zurückführen, sondern sie spiegeln den “resting state” Zustand (“eine allgemeine Situation”) im Gehirn wieder. Das neu entwickelte mathematische Verfahren der Eigenvektor Zentralität, das in abgewandelter Form auch bei Internet-Suchmaschinen zum Einsatz kommt, kann hier für die Auswertung verwendet werden: Das Gehirn kann man analog zum Internet als eine Menge verschiedener Netzwerke auffassen. Damit kann man die Methode der Eigenvektor Zentralität auch hier verwenden, um fMRT-Daten des menschlichen Gehirns besser und effizienter auswerten zu können.

Hier nutzt man die Tatsache, dass das Gehirn, genau wie das Internet, ein Netzwerk mit „small world properties“ (siehe [34]) ist, dass heißt, jeder Messpunkt im Gehirn oder jede Internetseite kann als ein Knotenpunkt in diesem Netzwerk angesehen werden. Die Knoten

können miteinander durch Kanten verbunden sein, genauso wie Homepages untereinander verlinkt sein können. Bei der Eigenvektor Zentralität werden die Knotenpunkte eines Graphen nach der Art und der Qualität der Verbindungen zu anderen Knoten gewichtet. Einerseits ist es wichtig, wie viele Verknüpfungen von einem Knoten selbst ausgehen, auch bezeichnet als Knotengrad, andererseits, wie stark zusammenhängend das Netzwerk der Nachbarknoten ist. Nach diesem Prinzip erfolgt die Reihenfolge der Internetseiten in der Suchliste, die mit hochfrequentierten Homepages verknüpft sind. Ein niedriger Rang entspricht dann einer guten Verknüpfung solcher Seiten. Die Vorteile der Auswertung von fMRT-Ergebnissen mit Hilfe der Eigenvektor Zentralität liegen darin, dass der Rechenaufwand überschaubar ist, sowie dass das Verfahren die Zusammenarbeit der Hirnregionen insgesamt betrachtet. Damit ist es zur Aufspürung für zustandsbedingte Aktivitätsmuster im Gehirn geeignet.

3.1 Zentralitäten in Graphen

Es gibt verschieden Möglichkeiten, die Knoten in einem Graphen ausgehend von ihrer strukturellen Lage angemessen zu bewerten. Hierfür definiert man Bewertungsmaße, welche beschreiben, wie sehr die von den Knoten repräsentierten Objekte im Graphen vernetzt sind. Diese Bewertungen sind die sogenannten Zentralitäten. Da diese auch von der Größe der betrachteten Graphen abhängen, werden die Zentralitäten normiert. Dadurch wird es möglich Zentralitäten von Graphen unterschiedlicher Größe zu vergleichen.

Das Gehirn soll so auf ein Modell abgebildet werden, dass die strukturellen Eigenschaften erhalten bleiben. Geeignet ist hierfür ein Graph $G = (V, E)$, bestehend aus einer Knotenmenge V und einer Kantenmenge E . Die Zentralitäten, auch Zentralitätsmaß genannt, ordnen dann jedem Knoten (Objekte), eine reelle Zahl zu.

Definition 1. Zentralitätsmaß

Ein Zentralitätsmaß Z ist eine Abbildung

$$Z : V \rightarrow \mathbb{R}_0, v_i \rightarrow Z(v_i), i = 1, \dots, n.$$

Dabei werden nur strukturelle Eigenschaften des Graphen berücksichtigt.

Definition 2. Zentralitätsvektor

Einen Vektor $Z(G)$, dessen Komponenten die Zentralitätswerte sämtlicher Knoten ent-

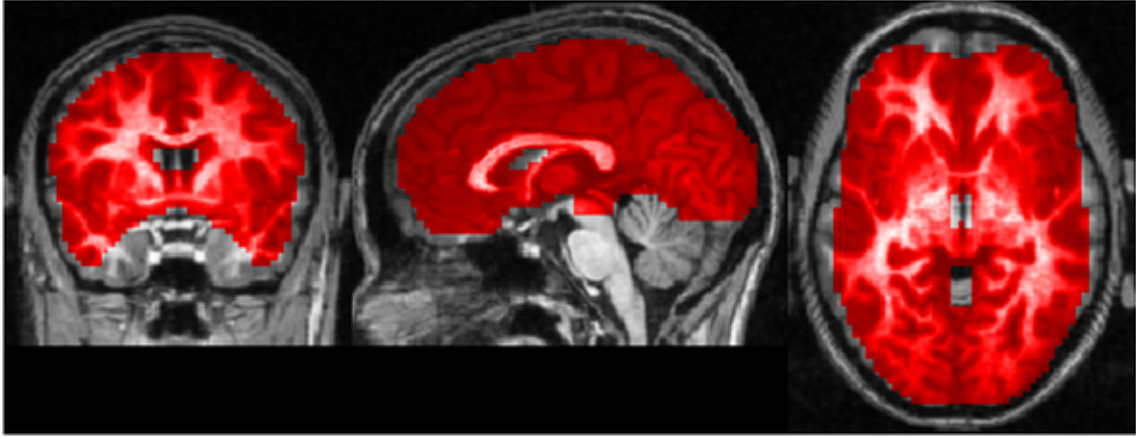


Figure 3.2.1: Die “region of interest”: Die Maske beinhaltet ≈ 40000 Voxels. Die Talairach Koordinaten für die Schnittposition sind $(0, 0, 0)$. [19]

halten bezeichnet man als Zentralitätsvektor. [2]

$$Z(G) := (Z(v_1), \dots, Z(v_n))$$

3.2 Berechnung der Eigenvektor Zentralität

Die Definition zentraler Knoten von Graphen ist in der Literatur nicht einheitlich. Die Gemeinsamkeit aller Definitionen besteht in einer symmetrischen Matrix, welche paarweise Ähnlichkeitsmaße beinhaltet. Im Sinne der Graphentheorie spricht man von einer modifizierten Adjazenzmatrix.

Sei A eine $n \times n$ Ähnlichkeitmatrix, wobei die Einträge $a_{ij}, i, j \in \{1, \dots, n\}$ ein paarweises Ähnlichkeitsmaß zwischen den fMRT-Zeitreihen im Voxel i und j beinhalten. Die Anzahl der Voxel n ist festgelegt durch den user-specified regions of interest (ROI). Das ist der Bereich, welcher vom Nutzer für die Auswertung der fMRT-Zeitreihen ausgewählt wird. [19] In den vorliegenden Experimenten ist der ROI das gesamte Gehirn unter Ausschluss vom Kleinhirn und besteht aus ca. 40000 Voxels.

Wie schon erwähnt kann jedes Voxel als Knoten in einem ungerichteten gewichteten Graphen betrachtet werden. Zusätzlich ist für jedes Voxel die räumliche Lage (x, y, z) im Hirn bekannt (MNI Koordinaten). Die Werte in der Ähnlichkeitsmatrix A entsprechen

den Kantengewichten. In graphentheoretisch basierten Anwendungen stellen die Gewichte die Distanzen zwischen den Knoten dar, welche nicht negativ sind. Folglich sind auch die Zentralitätsmaße nicht negativ. Diese Aussage wurde in [5] diskutiert.

Die Eigenvektor Zentralität favorisiert Knoten, welche zu jedem anderen Knoten hoch korreliert sind. Für den Knoten i ist die Eigenvektor Zentralität x_i definiert als der i -ten Eintrag im normalisierten Eigenvektor bezüglich des größten Eigenwertes der symmetrischen Ähnlichkeitsmatrix A . Für den größten Eigenwert λ bestimmt man den entsprechenden Eigenvektor x mit

$$Ax = \lambda x$$

oder äquivalent

$$x = \frac{1}{\lambda} Ax,$$

und

$$x_i = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} x_j$$

für $i = 1, \dots, n$, mit Proportionalitätsfaktor $\frac{1}{\lambda}$, so dass x_i proportional zur Summe der Ähnlichkeitswerte zu allen anderen Verbindungsknoten ist. Die Eindeutigkeit kann durch den Satz von Perron-Frobenius sichergestellt werden, welcher aussagt, dass jede quadratische Matrix mit streng positiven Einträgen $a_{ij} > 0$, $i, j = 1, \dots, n$ einen eindeutig größten reellen Eigenwert besitzt [13]. Das gilt ebenfalls für irreduzible quadratische Matrizen mit nicht negativen Einträgen. Eine irreduzible Matrix hat mindestens ein von Null verschiedenes nicht diagonales Element in jeder Zeile und Spalte.

Da angenommen wird, dass A die Distanzen zwischen den Knoten repräsentiert sind die Elemente $a_{ij} \geq 0$, $\forall i, j$. In diesem Kontext der fMRT-Analyse kann angenommen werden, dass A irreduzible ist, da die fMRT-Zeitreihen nicht vollständig unkorreliert sind, so dass eine Anzahl von Null verschiedenen Einträgen existiert. Demzufolge existiert ein Eigenvektor bezüglich des größten normalisierten Eigenwertes und die Einträge x_i liefern für jeden Knoten i ein Zentralitätsmaß, welches eindeutig definiert ist und nicht negativ ist. Symmetrische Matrizen mit negativen Einträgen können mehrere größte Eigenwerte besitzen, welche sich nicht unterscheiden. Deshalb ist die Bedingung der Nicht-Negativität wesentlich für die Eindeutigkeit der Eigenwerte.

Treten in der Ähnlichkeitsmatrix A negative Werte auf, so müssen die Werte a_{ij} neu



Abb. 3.2.2: Paarweise t-Test zwischen satt und hungrigen Probanden mit linearer Korrelation für die Eigenvektor Zentralität:
Ergebnis mit $p < 0.05$. Die Zentralitätswerte der Hirnregion Precuneus sind signifikant höher im hungrigen Zustand. Andere Regionen zeigen keine signifikanten Effekte. Die Talairach-Koordinaten in der Schnittposition sind $(-2, -50, 56)$. [19]

skaliert werden damit gilt $a_{ij} \geq 0, \forall i, j$. Die Art der Skalierung hängt vom Wertebereich der $a_{ij}, \forall i, j$ ab. [19]

Wenn die Berechnung der Eigenvektor Zentralität auf der gewöhnlichen Korrelationsmatrix beruht und der Datensatz, wie in 2.2 beschrieben, verwendet wird, dann ergibt sich beim Vergleich der Eigenvektor-Zentralitätswerte von satt und hungrigen Probanden die Abbildung 3.2.2. Diese Auswertungsmethode zeigt, dass bei hungrigen Probanden in der Hirnregion Precuneus signifikant höher Zentralitätswerte entstehen. [19]

4 Vektorquantisierung

4.1 Einleitung

Die Vektorquantisierung ist ein Verfahren zur Kompression von Datensätzen.[31] Bezüglich der fMRT-Zeitreihen gibt es zwei Motivationspunkte für die Anwendung einer Vektorquantisierung.

Zum einen sind die räumlich nächsten benachbarten Zeitreihen sehr ähnlich. Das bedeutet, dass der Unterschied zwischen den Zeitreihen könnte hauptsächlich Rauschen beeinhalteten. Folglich wird bei bestimmten Berechnungen, wie der partiellen Korrelation, siehe Kapitel 6, nur dieses Rauschen untersucht, wenn diese direkt benachbarten Zeitreihen in die Berechnung einfließen. Die partielle Korrelation zwischen direkt benachbarten Zeitreihen ist damit nicht sinnvoll. Demnach benötigt man einen bestimmten Radius ε für spezielle Untersuchungsmethoden. Das heißt, wenn die Zeitreihe x in einer Berechnung verwendet wird, dann sollen die räumlich nächsten Nachbarn mit einem Voxelabstand kleiner als ε von der Berechnung ausgeschlossen werden, aufgrund der Ähnlichkeit.

Zum anderen ermöglicht eine Vektorquantisierung eine Kompression der Daten der Art, dass bestimmte Verfahren, beispielsweise die partielle Korrelation, auf den großen Datensätzen, bestehend aus ca. 40 000 fMRT-Zeitreihen, ausführbar sind.

Grob formuliert ist die Vektorquantisierung ein Verfahren, um Mengen von Vektoren durch Repräsentanten zu approximieren. Ausgangspunkt hierfür ist eine Trainingssequenz, aus der ein Codebuch erzeugt wird. Die Einträge in diesem Codebuch heißen Codebuchvektoren, Prototypvektoren oder Referenzvektoren.[31]

Die Prototypvektoren bilden also in unserer Anwendung Repräsentanten für die fMRT-Zeitreihen, die durch ein Abstandsmaß eine “beste ” Approximation liefern. Ziel der Vektorquantisierung ist also, eine möglichst gute Repräsentation der Merkmalsvektoren (fMRT-Zeitreihen) zu erreichen. Die so resultierenden Prototypen haben dann den ge-

wünschten ε Radius. Dieser ε Radius entsteht allerdings nicht im original Raum \mathbb{R}^n (den Voxel-Koordinaten) des Datensatzes, sondern durch Anwendung der Vektorquantisierung im \mathbb{R}^m , wobei m die Zeitreihenlänge ist.

4.2 Prinzip

Vektorquantisierungstechniken kodieren eine Menge von Datenpunkten durch einen endlichen Satz von Referenzvektoren $v_i \in \mathbb{R}^D$, $i = 1, \dots, n$, welcher viel kleiner ist als die Anzahl der Datenpunkte. Die Kodierung eines Datenpunktes $v \in V$ erfolgt durch denjenigen Referenzvektor $w_{i(v)}$, für den der Beschreibungsfehler $d(v, w_{i(v)})$ am geringsten ist. Dabei ist d ein Ähnlichkeitsmaß, oft der euklidische Abstand. Dadurch wird die Menge V in eine Menge von Voronoi Zellen

$$V_i = \{v \in V \mid \|v - w_i\| \leq \|v - w_j\| \forall j\} \quad (4.2.1)$$

aufgeteilt. Sei $P(v)$ die Dichteverteilung der Datenpunkte über der Menge V , so ist der durchschnittliche Beschreibungsfehler aufgrund dieser Aufteilung in Voronoi Zellen durch

$$E = \int P(v) (v - w_{i(v)})^2 dv \quad (4.2.2)$$

gegeben. Das Anliegen besteht nun darin, diesen Beschreibungsfehler durch eine optimale Verteilung der Referenzvektoren w_i zu minimieren. Dafür gibt es verschiedene Ansätze und folglich unterschiedliche Adaptionsschritte. [22]

4.3 Der Neural Gas Algorithmus

Eine robuste Datenrepräsentation kann mit dem Neural Gas Algorithmus von Thomas Martinetz [22], welches die wesentliche Literatur für diesen Abschnitt ist, gefunden werden. Der stochastische Algorithmus ist insensitiv bezogen auf die Initialisierung und hat den Vorteil, dass er keine a priori Information über die Verteilung der Eingabesignale benötigt, welche nicht immer vorhanden sind.

Der Algorithmus entspricht der Diffusion von Molekülen eines Gases mit zunehmender Viskosität. Die Gaspartikel entsprechen den Referenzvektoren, die Datendichte einem

Potentialfeld. Bezugnehmend auf biologische Modelle werden in diesem Algorithmus die Partikel (Referenzvektoren) auch als Neurone bezeichnet. Seien die Referenzvektoren (Neurone) mit w_i bezeichnet, $i = 1, \dots, N$. Die Adjustierung der Referenzvektoren erfolgt mit der "soft-max"-Adaptionsregel. Dabei werden die impliziten Nachbarschaftsrelationen der Referenzvektoren im Eingabesignalraum zur Adjustierung der w_i verwendet. [33] Der Adaptionsschritt wird durch den Distanzrang bezüglich des Ähnlichkeitsmaßes d des Referenzvektors zum Eingabesignal bestimmt. Genauer, mit jedem präsentierten Eingabesignal v bestimmt man die Distanz oder die Nachbarschaftsrankfolge $w_{i_0}, w_{i_1}, \dots, w_{i_{N-1}}$ der Referenzvektoren zum Eingabesignal v . Hierbei ist w_{i_0} der zum Eingabesignal v nächste Referenzvektor, w_{i_1} der zweitnächste Referenzvektor zu v und w_{i_l} , $l = 0, \dots, N-1$ der Referenzvektor, für den l Vektoren w_j mit $\|v - w_j\| \leq \|v - w_{i_l}\|$ existieren. Sei $k_i(v, w)$ der Nachbarschaftsrank der dem Referenzvektor w_i zugeordnet ist. So gilt $k_i = 0$ für w_{i_0} , $k_i = 1$ für w_{i_1} und $k_i = l$ für w_{i_l} , $l = 0, \dots, N-1$. Der Nachbarschaftsrank $k_i(v, w)$ hängt neben v von allen Referenzvektoren $w = (w_1, \dots, w_N)$ ab. Im Sinne des Neuronenmodells berechnet $k_i(v, w)$ den Rang der Neuronen angeordnet gemäß ihren Abstand vom Trainingsmuster. Das nächstliegende Neuron wird also am meisten, das zweitnächste etwas weniger usw. adaptiert.

Der zur Adjustierung der Referenzvektoren verwandte Adaptionsschritt lautet ausgehend von der erhalten Nachbarschaftsrankfolge mit der Schrittweite $\epsilon \in [0, 1]$

$$\Delta w_i = \epsilon \cdot g_\lambda(k_i(v, w)) \cdot (v - w_i) \quad i = 1, \dots, N. \quad (4.3.1)$$

Die Schrittweite ϵ entspricht dem allgemeinen Ausmaß der Adjustierung. Die Funktion $g_\lambda(k_i(v, w))$ ist Eins für $k_i = 0$ und konvergiert mit steigendem k_i gegen Null mit einer Abklingkonstante λ , welche die Viskosität widerspiegelt. Typischerweise wählt man für die Funktion g eine Gaußfunktion, dass heißt $g_\lambda(k_i(v, w)) = e^{-k_i(v, w)/\lambda}$.

Betrachtet man einen Satz von Referenzvektoren $w = (w_1, \dots, w_N)$, so ist die mittlere Änderung $\langle \Delta w_i \rangle$ eines Referenzvektors mit Adaptionsschritt (4.3.1) durch die Mittelung von (4.3.1) über die Eingabesignalverteilung $P(v)$ gegeben und man erhält

$$\langle \Delta w_i \rangle = \epsilon \int P(v) g_\lambda(k_i(v, w)) (v - w_i) dv. \quad (4.3.2)$$

Die Gleichung (4.3.2) entspricht dem Gradienten der Kostenfunktion

$$E_{ng}(w, \lambda) = \frac{1}{2C(\lambda)} \sum_{i=1}^N \int P(v) g_{\lambda}(k_i(v, w)) (v - w_i)^2. \quad (4.3.3)$$

Dabei ist $C(\lambda)^{-1} = \sum_{k=0}^{N-1} g_{\lambda}(k)$ ein Normalisierungsfaktor, der nur von λ abhängt. E_{ng} kann als derjenige Beschreibungsfehler interpretiert werden, der minimiert werden muss, falls ein Datenpunkt v nicht nur durch den am nächsten liegenden Referenzvektor w_{i_0} beschrieben wird, sondern durch ein gewichtetes Mittel über alle Referenzvektoren w_i , mit $g_{\lambda}(k_i)/C(\lambda)$ als Gewichtungsfaktoren. Diese Gewichtungsfaktoren sind durch den jeweiligen Nachbarschaftsrang eines Referenzvektors w_i zum Datenpunkt v gegeben. Der Parameter λ beschreibt die Nachbarschaftsreichweite eines Adaptionsschrittes. Damit kann man die Gestalt der Kostenfunktion modulieren. Beispielsweise wird E_{ng} parabolisch für $\lambda \rightarrow \infty$. Wogegen für $\lambda \rightarrow 0$ die Kostenfunktion E_{ng} in den zu minimierenden Beschreibungsfehler E in (4.2.2) sehr gut approximiert wird. Eine Äquivalenz konnte bisher nicht gezeigt werden. [4]

Es soll nun ein kurzer Programmablauf angegeben werden. Dabei ist die Funktion g eine Gaußfunktion $g(\sigma, d) = e^{\left(\frac{-d^2}{2\sigma^2}\right)}$.

Algorithmus 1. *Neural Gas*

Initialisierung

wiederhole

wähle v_i

berechne $\|v_i - w_j\|$

bestimme $k_j(v_i, w) = |\{k \mid \|v_i - w_k\| \leq \|v_i - w_{i_0}\||$

adaptiere $\Delta w_j = \epsilon \cdot g_{\lambda}(\sigma, k_j(v_i, w)) \cdot (v_i - w_j)$

$\epsilon = \epsilon \cdot \alpha_1; \sigma = \sigma \cdot \alpha_2$

bis Konvergenz

Das Training adaptiert in jedem Schritt alle Neuronen. Typischerweise ist $\alpha_1 = \alpha_2 \approx 0.99$, dass heißt, die Nachbarschaft und die Lernrate werden nach jedem Schritt verringert. Die Zuordnung der Datenpunkte zu den Referenzvektoren erfolgt über die Winner-Takes-All Funktion. In dieser Funktion wird ein Datenpunkt auf das nächst gelegene Neuron abgebildet oder anders formuliert der Datenpunkt wird dem Referenzvektor

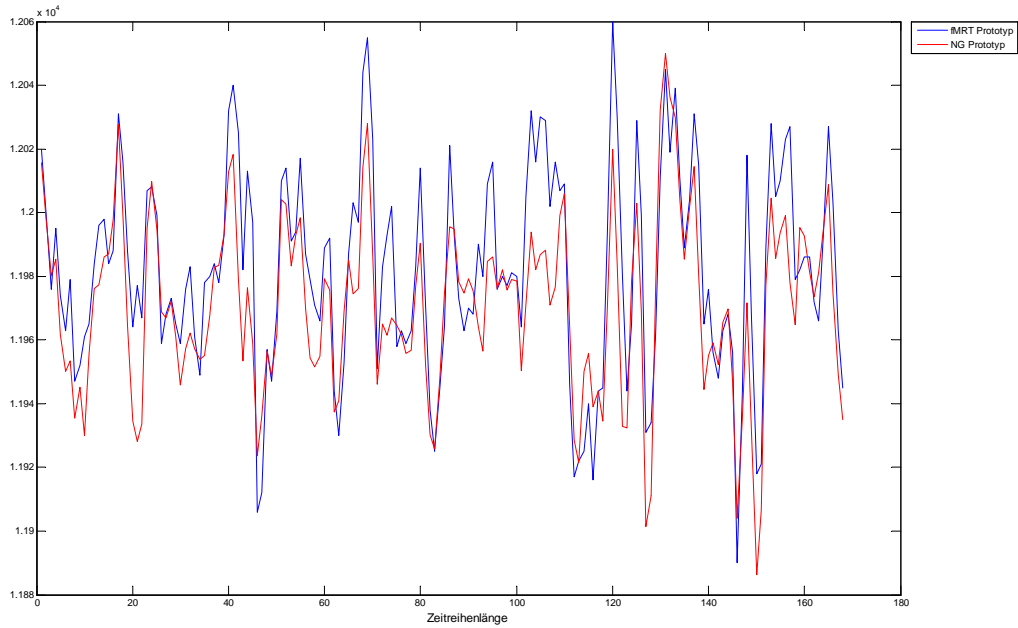


Figure 4.3.1: Darstellung des Glättungseffektes entstehend durch den Neural Gas Algorithmus:

Die blaue Zeitreihe ist der Prototyp, welcher vom Neural Gas Algorithmus ermittelt wurde und die rote ist die für diesen Prototypen am besten approximierte Zeitreihe.

zugeordnet, für den die Distanz am kleinsten ist. Die Winner-Takes-All Funktion ist eine charakteristische Funktion der Form

$$1_{WTA}^j = \begin{cases} 1, & \text{falls } d(v, w_j) \leq d(v, w_i), \text{ für } i = 1, \dots, N \\ 0, & \text{sonst} \end{cases}.$$

Im Ergebnis gilt $w_i = \frac{1}{\#\{v \in V_i\}}$ für die Menge V (siehe (4.2.1)) und für die Dichte $\varrho(w_i)$ der Prototypen gilt $\varrho(w_i) \sim P(w_i)^\alpha$, wobei $\alpha = \frac{\dim}{\dim+2}$ und \dim die intrinsische Dimension nach Hausdorff ist.[28]

Die Referenzvektoren w_i entsprechen also Mittelungen der Datenvektoren in den zugeordneten Voronoi Zellen V_i , was zu einem Glättungseffekt führt. Dieser stellt jedoch einen Informationsverlust dar. Dieser Effekt ist in Abbildung 4.3.1 dargestellt. Zur Vermeidung des Glättungseffektes werden anstatt der Referenzvektoren resultierend aus dem Neural

Gas Algorithmus, originale fMRT-Zeitreihen verwendet, die am besten auf die Referenzvektoren abgebildet werden. Dieses Vorgehen entspricht auch dem Median Neural Gas Algorithmus. Die resultierenden Referenzvektoren sind also originale fMRT-Zeitreihen.

Die Anzahl der Referenzvektoren im Neural Gas Algorithmus für das vorliegende Problem richtet sich nach der biologischen Struktur des Gehirns und wird auf 100 festgelegt. Diese Anzahl entspricht in etwa der Anzahl der Hirnareale, welches genauer in [32] nachgelesen werden kann. Weiter besteht die Annahme, dass sich ähnliche Prototypen in Cluster vereinigen lassen. Dafür verwendet man einen speziellen Clusteralgorithmus namens Affinity Propagation.

5 Clusterung der fMRI Prototypen

Um eine weitere Datenreduktion zu erreichen, werden die Prototypen aus der Neural Gas Vektorquantisierung geclustert. Diese Cluster sollten kompakt und gut separierbar sein. In diesem Kapitel basiert die Theorie, falls nicht anders angegeben, auf der Veröffentlichung von Frey und Dueck [11].

5.1 Einleitung

Eine Methode zur Clusterung von Daten ist Affinity Propagation. Für diese müssen paarweisen Ähnlichkeiten der Datenobjekte bekannt sein. Affinity Propagation basiert auf der Methode des “message passing” als Realisierung des sogenannten Max-Sum-Algorithmus für Faktorgraphen. Dabei wird der Faktorgraph aus der zugrundeliegenden Kostenfunktion des Affinity Propagation Algorithmus gebildet. Eine detaillierte Beschreibung ist in [11] zu finden.

In Affinity Propagation wird jeder Datenvektor (Datenobjekt) als potentiell Clusterzentrum (Exemplar) angesehen. Iterativ werden sogenannte Botschaften zwischen den Datenobjekten ausgetauscht. Diese beschreiben, wie attraktiv ein Datenobjekt von anderen Datenobjekten als potentiell Clusterzentrum gesehen wird, beziehungsweise, wie attraktiv ein Datenobjekt sich selbst “sieht”. Diesen Vorgang bezeichnet man auch als “message passing”.

Der Algorithmus bricht bei gegebener Parametereinstellung (Präferenzwert) mit der Ausgabe der Exemplare (Datenobjekte, die die Clusterzentren darstellen) ab, sowie den Zuordnungen der Datenpunkte zu den Clusterzentren. Die Einstellung der Präferenzwertes bestimmt maßgeblich die Anzahl der Cluster. Stabile Clusterzahlen sind solche, die sich für gewisse Parameterbereiche des Präferenzwertes nicht ändern. Nur stabile Konfigurationen sind stabile Clusterlösungen. Im Beispiel 5.2.1 erkennt man, dass für

die zugrundeliegenden fMRT-Zeitreihen die Präferenzbereiche $[1, 8; 2, 8]$, $[3, 3; 4, 8]$ und $[5, 1; 8, 4]$ stabile Clusterlösungen auftreten.

5.2 Anwendung auf die fMRT-Prototypen

Beim Affinity Propagation Algorithmus wird jedem Prototyp crisp ein Cluster zugeordnet, dass heißt, jeder Prototyp gehört zu genau einem Cluster. Die erste stabile Clusterlösung wird für den Präferenzwert 2, 5, siehe Abbildung 5.2.1. Das entspricht einer stabilen Lösung mit maximaler Clusteranzahl, in diesem Fall 7 Cluster. Die Clusterlösung und die entsprechenden Exemplare sind in Abbildung 5.2.2 abgebildet, wobei unterschiedlich gefärbte Zeitreihen verschiedene Cluster darstellen. Die schwarzen Zeitreihen sind die Exemplare, welche genau mittig in jedem Cluster liegen und dem Clusterzentrum entsprechen. Man sagt dazu auch Median-Clusterung.

Da die Entstehungspunkte der fMRT-Zeitreihen durch die Zuordnung der Voxel bekannt sind, kann man nun die Clusterverteilung im Gehirn unter Ausnutzung der Koordinaten der Voxel zurückverfolgen. Die Abbildung 5.2.3 zeigt die Clusterbildung im gesamten Gehirn. Laut dieser Abbildung kann angenommen werden, dass die Cluster kompakt und gut separiert sind. Die Betrachtung der Clusterlösung in der XY-Ebene, XZ-Ebene und YZ-Ebene, siehe Abbildung 5.2.4, 5.2.5 und 5.2.6 bestätigt die Annahme. Für eine detaillierte Analyse und Interpretation der Clusterlösung sollten spezielle Validierungsmaße herangezogen werden. Diese würden jedoch den Rahmen dieser Arbeit sprengen und sollten später verfolgt werden.

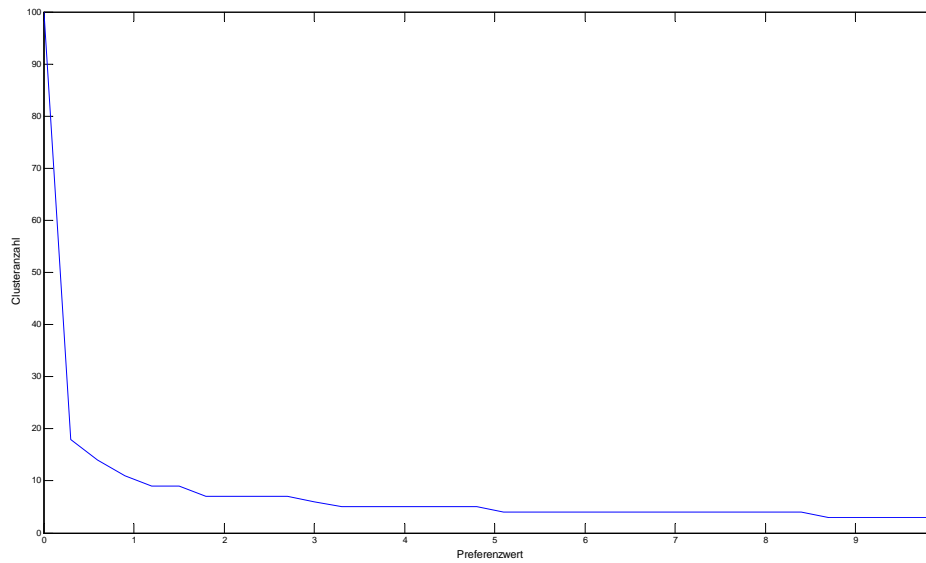


Figure 5.2.1: Präferenzwerte für stabile Phasen

Dieser Wert kann auf der abszissen Achse abgelesen werden und die ordinate Achse gibt die Clusteranzahl an.

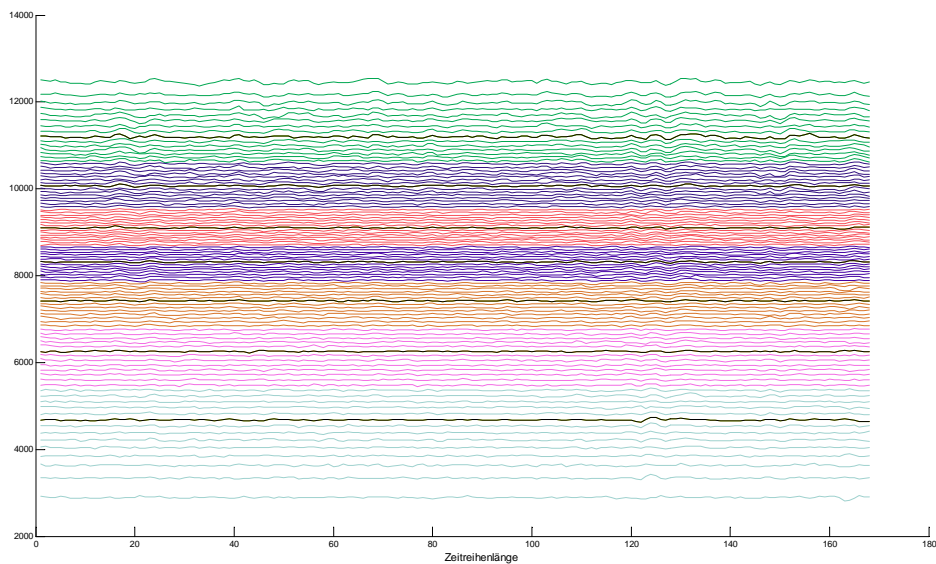


Figure 5.2.2: Geclusterten fMRT-Prototypen mit den Exemplaren.

Jede Farbe stellt ein Cluster dar und die schwarzen Zeitreihen die jeweiligen Clusterzentrum.

5 Clusterung der fMRI Prototypen

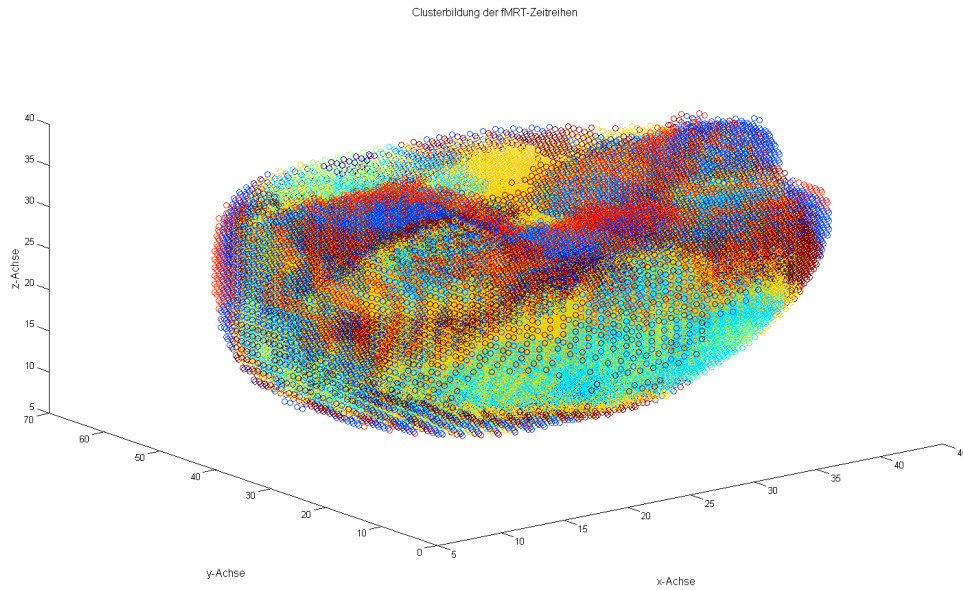


Figure 5.2.3: Clusterverteilung im vollständigen Gehirn, gemäß der Clusterlösung und der Rückabbildung bezüglich der Voxelkoordinaten.

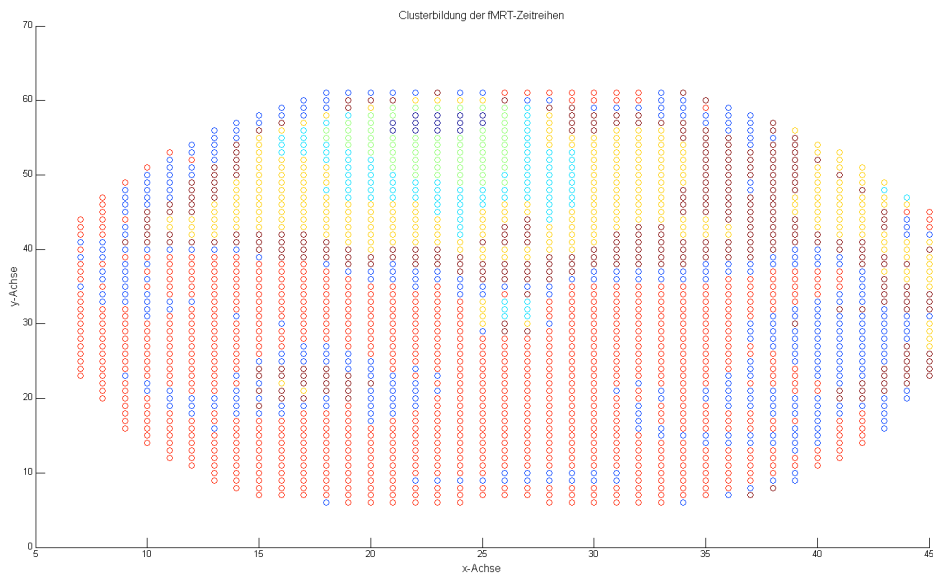


Figure 5.2.4: Clusterbildung in der XY-Ebene.

5 Clusterung der fMRI Prototypen

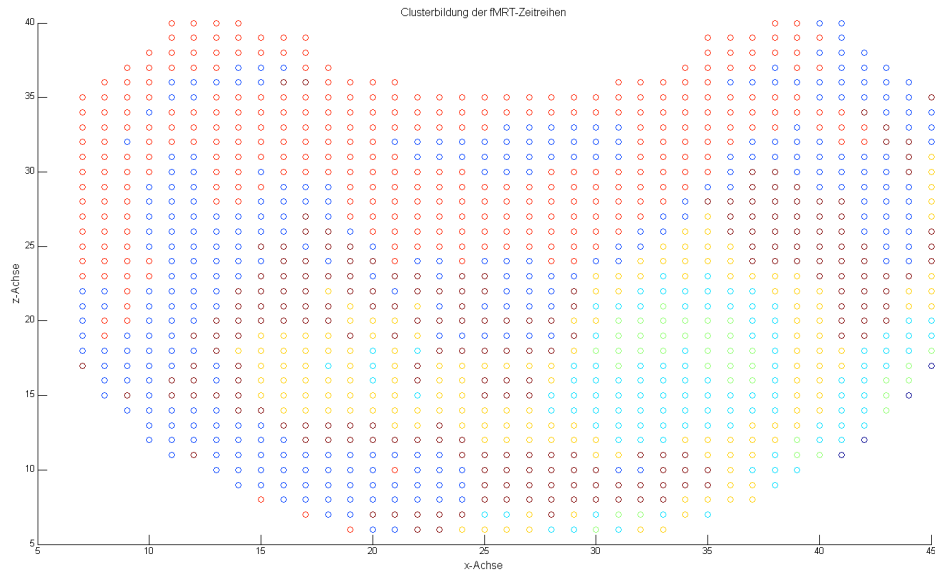


Figure 5.2.5: Clusterbildung in der XZ-Ebene.

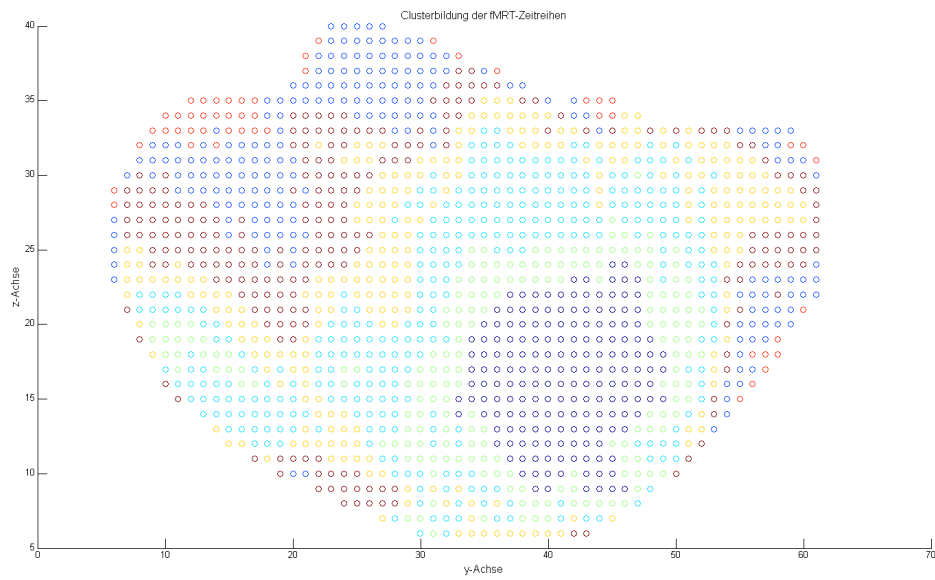


Figure 5.2.6: Clusterbildung in der YZ-Ebene

6 Partielle Korrelationen

Eines der Hauptanliegen der statistischen Analyse von Daten ist das Erkennen von Abhängigkeiten und Zusammenhängen verschiedener Zufallsgrößen. Die Kovarianz und der Korrelationskoeffizient sind Maße für die lineare Abhängigkeit von Zufallsgrößen, welche in dieser Arbeit die fMRT-Zeitreihen darstellen [30].

6.1 Der Korrelationskoeffizient

Für die theoretischen Grundlagen dieses Abschnitts bezieht man sich auf die Bücher [14] und [26].

Der Einfachheit halber werden zunächst drei reelle Zufallsvariablen X, Y und Z betrachtet und die entsprechenden Erwartungswerte mit $\mu_X = E(X)$, $\mu_Y = E(Y)$, $\mu_Z = E(Z)$ bezeichnet. Die Varianzen seien $VAR(X) = D^2X$, $VAR(Y) = D^2Y$, $VAR(Z) = D^2Z$ und die Kovarianzen $\sigma_{XY} = COV(X, Y)$, $\sigma_{XZ} = COV(X, Z)$ und $\sigma_{YZ} = COV(Y, Z)$. Für den Vektor (X, Y, Z) ergibt sich damit die Kovarianzmatrix

$$\Sigma = COV \left(\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \right) = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{XZ} \\ \sigma_{YX} & \sigma_Y^2 & \sigma_{YZ} \\ \sigma_{ZX} & \sigma_{ZY} & \sigma_Z^2 \end{pmatrix}.$$

Sei $X = (x, y)^T$, $x, y \in \mathbb{R}$ ein Zufallsvektor, wobei die Varianzen $D^2x = \sigma_x^2$ und $D^2y = \sigma_y^2$ existieren. Dann ist die Kovarianz wie folgt definiert

$$COV(X, Y) := E((X - EX)(Y - EY)).$$

Mit der Kovarianz kann nun der Korrelationskoeffizient, auch marginale Korrelation

genannt, von X und Y definiert werden

$$CORR(X, Y) = \frac{COV(X, Y)}{\sigma_x \sigma_y}$$

und es gilt immer $-1 \leq CORR(X, Y) \leq 1$. Dies folgt aus der schwarzischen Ungleichung. [25]

Der Begriff marginale Korrelation bezieht sich darauf, dass dazu nur die gemeinsame Verteilung der Zufallsvariablen X und Y verwendet wird. Die Korrelation $CORR(X, Y) = \pm 1$, wenn X und Y exakt linear abhängig sind. Je höher der Betrag von $CORR(X, Y)$, desto höher ist die lineare Abhängigkeit zwischen X und Y . Falls $CORR(X, Y) = 0$ ist, so heißen X und Y unkorreliert. Wenn X und Y unabhängig sind, dann folgt, dass $CORR(X, Y) = 0$. Umgekehrt kann im Allgemeinen nicht aus der Unkorreliertheit die Unabhängigkeit gefolgert werden. Eine Ausnahme ist, wenn die Zufallsvariablen multivariat normalverteilt sind, dann gilt auch die Umkehrung. Daher ist im Normalverteilungsfall die Korrelation ein Maß für die Stärke der Abhängigkeit von Zufallsvariablen. [6]

Es gibt verschiedene Möglichkeiten den Korrelationskoeffizienten für normalverteilte Zufallsvariablen zu schätzen.

6.1.1 Korrelation nach Pearson

Es seien die Zufallsvariablen $X = (x_1, \dots, x_n)$ und $Y = (y_1, \dots, y_n)$ in ihrer Grundgesamtheit normalverteilt. Ein Schätzer für die Korrelation $CORR(X, Y)$ ist der empirische Korrelationskoeffizient nach Pearson, mit r_{XY} bezeichnet.

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{j=1}^n (y_j - \bar{y})^2}}, \quad (6.1.1)$$

wobei $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ und $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ ist.

Wenn eine endliche Menge von Zufallsvariablen X_1, \dots, X_n gegeben ist, dann können die

Korrelationskoeffizienten in einer sogenannten Korrelationsmatrix

$$A = \begin{pmatrix} r_{x_1x_1} & r_{x_1x_2} & \cdots & r_{x_1x_n} \\ r_{x_2x_1} & r_{x_2x_2} & & \\ \vdots & & \ddots & \\ r_{x_nx_1} & & & r_{x_nx_n} \end{pmatrix}$$

angegeben werden. Diese Matrix ist symmetrisch und quadratisch. Weiter stehen auf der Hauptdiagonalen nur Einsen, da jede Zufallsvariable zu sich selbst korreliert ist.[14]

Die bisherige Grundlage der Berechnung der Eigenvektor Zentralität ist eine skalierte Korrelationsmatrix, welche die Ähnlichkeitsmatrix darstellt. Eine Skalierung der Korrelationsmatrix wird vorgenommen, damit die Voraussetzung für die Anwendung der Eigenvektorzentralität erfüllt ist. Die Voraussetzung besteht darin, dass die Ähnlichkeitsmatrix nur positive Elemente enthält (siehe Aussage des Satzes von Perron-Frobenius in Kapitel 3). Nach einer Skalierung der Korrelationsmatrix A der Form $\tilde{A} = A + 1$ hat diese Matrix dann nur noch positive Elemente, da gilt $-1 \leq a_{ij} \leq 1$, $\forall i, j$. Diese Matrix \tilde{A} ist die bisherige Grundlage für die Berechnung der Eigenvektor Zentralität, welche eine Größe von $\approx 39000 \times 39000$ hat und fast nur hoch korrelierte Werte enthält. Auf Grund der Vektorquantisierung mit dem Neural Gas Algorithmus und der anschließenden Clusterung der Prototypen mit Affinity Propagation kann die Korrelationsmatrix auf Basis der Exemplare bestimmt werden. Diese Matrix $A_{Exemplare}$ ist nun wesentlich kleiner und hat in dem der Arbeit zu grundlegenden Datensatz die Größe 7×7 , da sieben Exemplare aus Affinity Propagation resultieren (siehe Kapitel 5). Es ist zu erkennen, dass die Matrix $A_{Exemplare}$ keine hoch korrelierten Werte beinhaltet. Ein Grund dafür ist, dass jeweils ähnliche Zeitreihen in einem Cluster zusammengefasst sind, dass heißt die Clusterzentren (Exemplare) sind relativ unterschiedlich. Somit kann es kaum noch hochkorrelierte Werte geben.

$$A_{Exemplare} = \begin{pmatrix} 1 & 0.3058 & 0.2246 & 0.1975 & 0.0877 & 0.1962 & 0.3794 \\ 0.3058 & 1 & 0.6138 & 0.4150 & 0.2776 & 0.2592 & 0.5262 \\ 0.2246 & 0.6138 & 1 & 0.4349 & 0.4899 & 0.2401 & 0.5137 \\ 0.1975 & 0.4150 & 0.4349 & 1 & 0.3568 & 0.2580 & 0.4670 \\ 0.0877 & 0.2776 & 0.4899 & 0.3568 & 1 & 0.3623 & 0.3317 \\ 0.1962 & 0.2592 & 0.2401 & 0.2580 & 0.3623 & 1 & 0.3227 \\ 0.3794 & 0.5262 & 0.5137 & 0.4670 & 0.3317 & 0.3227 & 1 \end{pmatrix} \quad (6.1.2)$$

6.1.2 Rangkorrelationskoeffizient nach Spearman

Weitere Schätzer für empirische Korrelationskoeffizienten sind Rangkorrelationskoeffizienten. Im Gegensatz zum Korrelationskoeffizienten von Pearson misst der Rangkorrelationskoeffizient nicht nur lineare Zusammenhänge, sondern erfasst den Zusammenhang nichtlinear.[14]

Bei ordinal skalierten Merkmalen (Variablen) wird die Stärke des Zusammenhangs gemessen an der Stärke der Übereinstimmung der Reihenfolge von den Realisierungen der Variablen (Merkmalsausprägungen).

Seien $X = (x_1, \dots, x_n)$ und $Y = (y_1, \dots, y_n)$ ordinal skalierte Zufallsvariablen, so kann deren Korrelation mittels des Spearmanschen Rangkorrelationskoeffizienten, mit r^s bezeichnet, geschätzt werden. Dabei werden nur die Ranginformationen zur Schätzung verwendet. Die n Realisierungen der Zufallsvariablen von X und Y werden aufsteigend oder absteigend geordnet und anschließend nummeriert. Aus der Nummerierung werden für die Realisierungen x_i und y_i ($i = 1, \dots, n$) die Rangzahlen R_{x_i} und R_{y_i} vergeben. So ergibt sich der Korrelationskoeffizient r^s aus der Schätzung nach Pearson, indem in Gleichung (6.1.1) die Rangzahlen R_{x_i} und R_{y_i} verwendet werden.[14]

$$\begin{aligned} r_{XY}^s &= \frac{\sum_{i=1}^n (R(x_i) - \overline{R(x)}) (R(y_i) - \overline{R(y)})}{\sqrt{\sum_{i=1}^n (R(x_i) - \overline{R(x)})^2 \cdot \sum_{i=1}^n (R(y_i) - \overline{R(y)})^2}} \\ &= 1 - \frac{6 \sum_{i=1}^n (R(x_i) - R(y_i))^2}{n(n^2 - 1)} \end{aligned}$$

Dieser Korrelationskoeffizient kann auch für metrisch skalierte Werte berechnet werden,

wie es bei den fMRT-Zeitreihen der Fall ist. Die Berechnung von r_{X_1, \dots, X_m}^s für m Exemplare ergibt folgende Korrelationsmatrix $A_{Exemplare}^s$

$$A_{Exemplare}^s = \begin{pmatrix} 1 & 0.2570 & 0.2110 & 0.2607 & 0.0886 & 0.1378 & 0.3574 \\ 0.2570 & 1 & 0.5935 & 0.3850 & 0.2394 & 0.2055 & 0.5116 \\ 0.2110 & 0.5935 & 1 & 0.4083 & 0.4413 & 0.1906 & 0.4214 \\ 0.2607 & 0.3850 & 0.4083 & 1 & 0.4413 & 0.1906 & 0.4214 \\ 0.0886 & 0.2394 & 0.4413 & 0.3231 & 1 & 0.3531 & 0.2510 \\ 0.1378 & 0.2055 & 0.1906 & 0.2367 & 0.3531 & 1 & 0.2793 \\ 0.3574 & 0.5116 & 0.4214 & 0.4465 & 0.2510 & 0.2793 & 1 \end{pmatrix}. \quad (6.1.3)$$

Der Vergleich zwischen den Matrizen $A_{Exemplare}$ (6.1.2) und $A_{Exemplare}^s$ (6.1.3) zeigt ähnliche Werte, obwohl die Schätzung mit einem Rangkorrelationskoeffizienten die nicht linearen Zusammenhänge erfasst. Dadurch besteht keine Notwendigkeit die Korrelation nach Spearman zu schätzen.

6.1.3 Rangkorrelationskoeffizient nach Kendall

Der Kendallsche Rangkorrelationskoeffizient ist eine weitere Methode zur Schätzung der Korrelation, welche ebenfalls nur auf den Rängen der Zufallsvariablen basiert. Für die Berechnung bilden die n Realisationen der Zufallsvariablen X und Y n Paare $(x_1, y_1), \dots, (x_n, y_n)$ und es werden wieder die Rangzahlen R_{x_i} und R_{y_j} bestimmt. Die Paare $(x_1, y_1), \dots, (x_n, y_n)$ werden so angeordnet, dass der Rang der Realisation von X aufsteigend geordnet ist. Dadurch wird auch eine Reihenfolge der Rangzahlen der Realisationen von Y festgelegt. In dieser Reihenfolge wird für jede Rangzahl R_{y_i} die Anzahl q_i der Rangzahlen R_{y_j} bestimmt, die kleiner oder gleich R_{y_i} ist. Der Kendallsche Korrelationskoeffizient ergibt sich dann zu

$$r^k = 1 - \frac{4 \sum_{i=1}^n q_i}{n(n-1)}.$$

[14] Damit erhält man die nachfolgende Korrelationsmatrix $A_{Example}^k$ mit der Schätzung der Korrelation nach Kendall r_{X_1, \dots, X_m}^k für m Exemplare

$$A_{Example}^k = \begin{pmatrix} 1 & 0.1747 & 0.1483 & 0.1835 & 0.0601 & 0.0898 & 0.2484 \\ 0.1747 & 1 & 0.4225 & 0.2621 & 0.1597 & 0.1415 & 0.3552 \\ 0.1483 & 0.4225 & 1 & 0.2767 & 0.3133 & 0.1297 & 0.2924 \\ 0.1835 & 0.2621 & 0.2767 & 1 & 0.2176 & 0.1657 & 0.3050 \\ 0.0601 & 0.1597 & 0.3133 & 0.2176 & 1 & 0.2376 & 0.1704 \\ 0.0898 & 0.1415 & 0.1297 & 0.1657 & 0.2376 & 1 & 0.1908 \\ 0.2484 & 0.3552 & 0.2924 & 0.3050 & 0.1704 & 0.1908 & 1 \end{pmatrix}. \quad (6.1.4)$$

Die Matrixelemente in $A_{Example}^k$ haben generell kleinere Werte als in der Matrix $A_{Example}$. Ein Vergleich zwischen den Matrizen $A_{Example}$ (6.1.2), $A_{Example}^s$ (6.1.3) und $A_{Example}^k$ (6.1.4) zeigt jedoch nur geringe Unterschiede. Deshalb kann die Schätzung der Korrelation weiterhin mit dem Pearsonschen Korrelationskoeffizient erfolgen.

6.1.4 Signifikanztests

Die Korrelationsmatrix ist die Basis für weitere Berechnungen und soll nur signifikant von Null verschiedene Werte enthalten. Um diese Hypothese zu überprüfen, ob eine Korrelation signifikant von einem Wert ϱ_0 abweicht, wird zunächst die Korrelation und ϱ_0 in Fischer Z-Werte transformiert.

Die empirischen Korrelationskoeffizienten nach Pearson sind nicht normalverteilt. Die Verteilung ist "eingipflig" nach rechts verzerrt, dass heißt linksschief. Vor der Berechnung der Prüfgröße für den Signifikanztest muss daher erst eine Korrektur der Verteilung mit Hilfe der Fisher-Transformation vorgenommen werden. Aus der Umformung der Korrelationskoeffizienten (mit Tabelle H im Anhang des Buches [6]) ergeben sich die transformierten Fisher Z-Werte, wobei Z für die berechnete Korrelation und Z_0 für den Wert ϱ_0 steht. Diese Fisher Z-Werte sind dann annähernd normalverteilt.

Sind die den Korrelationen zugrundeliegenden Variablen multivariat normalverteilt, folgt die Prüfgröße

$$z = \sqrt{n - 3 - (k - 2)} \cdot (Z - Z_0)$$

approximativ einer Standardnormalverteilung, wobei n die Anzahl der Realisierung der Zufallsvariablen und k die Anzahl der Zufallsvariablen ist. Für den marginalen

Korrelationskoeffizienten ist $k = 2$, somit ergibt sich die Prüfgröße

$$z = \sqrt{n - 4} \cdot (Z - Z_0).$$

Die Korrelation weicht bei einem zweiseitigen Test statistisch bedeutsam von ϱ_0 ab, wenn z außerhalb der Bereiche $-1,96 \leq z \leq 1,96$ liegt mit dem Signifikanzniveau $\alpha = 0,05$ oder $-2,58 \leq z \leq 2,58$ für $\alpha = 0,01$. [6]

6.2 Bedingte und Partielle Korrelationen

Dieser Abschnitt zur bedingten und partiellen Korrelation wurde hauptsächlich aus der Literatur [31] und [3] entnommen.

Unter einer bedingten Korrelation von X und Y unter $Z = z$, abgekürzt $r_{X,Y|Z=z}$, versteht man die Korrelation, die aus den bedingten Verteilungen von $(X, Y) | Z = z$ und $X | Z = z$, $Y | Z = z$ berechnet wird,

$$r_{X,Y|Z=z} = \frac{COV(X, Y | Z = z)}{\sqrt{VAR(X | Z = z)} \cdot \sqrt{VAR(Y | Z = z)}}. \quad (6.2.1)$$

Bei $r_{X,Y|Z=z}$ wird gemessen, wie stark der lineare Zusammenhang der bedingten Variablen X und Y für einen festgelegten Wert $Z = z$ ist. Die bedingten Verteilungen, die für die Bestimmung relevant sind, besitzen die Dichte

$$f_{XY|Z=z}(x, y) = \frac{f_{XYZ}(x, y, z)}{f_Z(z)},$$

$$f_{X|Z=z}(x) = \frac{f_{XZ}(x, z)}{f_Z(z)},$$

$$f_{Y|Z=z}(y) = \frac{f_{YZ}(y, z)}{f_Z(z)},$$

wobei von der Existenz der bedingten Dichten ausgegangen werden soll. Ein Spezialfall tritt auf, wenn alle Variablen gemeinsam normalverteilt sind, dass heißt $(X, Y, Z)^T \sim N(\mu, \Sigma)$. Dies wird im Allgemeinen auch für die fMRT-Zeitreihen angenommen. Da an dieser Stelle die bedingte Korrelation nicht von z abhängt, kann eine einfachere Bezeichnung $r_{X,Y|Z}$ statt $r_{X,Y|Z=z}$ gewählt werden. Die bedingte Korrelation ergibt sich

dann zu

$$r_{X,Y|Z} = \frac{COV(X, Y|Z)}{\sqrt{VAR(X|Z)} \cdot \sqrt{VAR(Y|Z)}}.$$

Bemerkenswert ist, dass im Allgemeinen gilt $r_{X,Y} \neq r_{X,Y|Z}$.

Die partielle Korrelation zielt darauf ab, zu bestimmen, inwieweit die Korrelation zwischen zwei Variablen X und Y auf den Einfluss einer dritten Variable Z zurückzuführen ist. Dabei wird nicht wie bei der bedingten Korrelation auf den Wert $Z = z$ bedingt, sondern versucht, den gesamten linearen Anteil von Z herauszurechnen. Im Normalverteilungsfall $(X, Y, Z)^T \sim N(\mu, \Sigma)$ gilt, dass die bedingte Korrelation $r_{X,Y|Z=z}$ gleich der partiellen Korrelation $r_{X,Y|Z}$ ist.

Dazu formuliert man die linearen Regressionsmodelle von X auf Z und Y auf Z

$$X = a_0 + a_1 Z + U_{XZ} \quad (6.2.2)$$

$$Y = b_0 + b_1 Z + U_{YZ}, \quad (6.2.3)$$

wobei X und Y die Regressanden (abhängige Variablen), Z der Regressor (unabhängige Variable), a_0, a_1, b_0, b_1 die theoretischen Regressionskoeffizienten und U_{XZ}, U_{YZ} die zufälligen Störvariablen sind, unter der Annahme, dass $U \sim N(0, \sigma^2)$. Die Parameter a_0, a_1 werden so bestimmt, dass

$$\widehat{X} = a_0 + a_1 Z \quad (6.2.4)$$

die beste lineare Prognose für die Gleichung (6.2.2) bildet. Die Gestaltung von a_0, a_1 ist der Art, dass die erwartete quadratische Abweichung zwischen X und \widehat{X} minimal wird

$$E \left\{ (X - a_0 - a_1 Z)^2 \right\} \xrightarrow{a_0, a_1} \min. \quad (6.2.5)$$

Auf gleicher Weise minimiert man die erwartete quadratische Abweichung zwischen Y und \widehat{Y}

$$E \left\{ (Y - b_0 - b_1 Z)^2 \right\} \xrightarrow{b_0, b_1} \min. \quad (6.2.6)$$

Die Existenz der Momente vorausgesetzt, lässt sich zeigen, dass die Koeffizienten vom ersten Minimierungsproblem (6.2.5) folgendes ergeben

$$a_1 = \frac{\sigma_{XZ}}{\sigma_Z^2}, \quad (6.2.7)$$

$$a_0 = \mu_X - a_1 \mu_Z. \quad (6.2.8)$$

Unter Verwendung der Methode der kleinsten Quadrate kann man $\widehat{a}_0, \widehat{a}_1$, die Schätzer von a_0, a_1 , ableiten. Damit ergibt sich das Minimierungsproblem

$$\sum_{i=1}^n u_i^2 = \sum_{i=1}^n (x_i - (\widehat{a}_0 + \widehat{a}_1 z_i))^2 \xrightarrow{\widehat{a}_0, \widehat{a}_1} \min,$$

$$\sum_{i=1}^n u_i^2 = \sum_{i=1}^n (x_i - (\widehat{x}_i))^2 \longrightarrow \min,$$

wobei $x_i, i = 1, \dots, n$ die beobachteten Werte der abhängigen Variable X , \widehat{x}_i die auf der Regressionsgeraden liegenden geschätzten Werte an der Stelle $z_i, i = 1, \dots, n$ und $u_i = x_i - \widehat{x}_i$ die nicht durch die Regressionsgerade erklärten Abweichungen sind. Die Abweichungen werden Residuen genannt. Das Minimierungsproblem wird durch die partiellen Ableitungen einer Funktion zweier Veränderlicher $\widehat{a}_0, \widehat{a}_1$ und der anschließenden Bestimmung der Nullstelle gelöst.

$$f(\widehat{a}_0, \widehat{a}_1) = \sum_{i=1}^n (x_i - \widehat{a}_0 + \widehat{a}_1 z_i)^2 \quad (6.2.9)$$

$$\frac{\partial f}{\partial \widehat{a}_0} = 2 \cdot \sum_{i=1}^n (x_i - \widehat{a}_0 + \widehat{a}_1 z_i) \cdot (-1) \stackrel{!}{=} 0 \quad (6.2.10)$$

$$\rightarrow \sum_{i=1}^n x_i - n\widehat{a}_0 - \widehat{a}_1 \sum_{i=1}^n z_i = 0 \quad (6.2.11)$$

$$\frac{\partial f}{\partial \widehat{a}_1} = 2 \cdot \sum_{i=1}^n (x_i - \widehat{a}_0 + \widehat{a}_1 z_i) \cdot (-z_i) \stackrel{!}{=} 0 \quad (6.2.12)$$

$$\rightarrow \sum_{i=1}^n x_i z_i - \widehat{a}_0 \sum_{i=1}^n z_i - \widehat{a}_1 \sum_{i=1}^n z_i^2 = 0 \quad (6.2.13)$$

Man erhält ein Gauß'sches Normalgleichungssystem, dass heißt ein lineares Gleichungssystem in den Variablen \widehat{a}_0 und \widehat{a}_1 . Aus Gleichung (6.2.11) ergibt sich der Koeffizient \widehat{a}_0 wie in Gleichung (6.2.8)

$$\widehat{a}_0 = \frac{-\widehat{a}_1 \sum_{i=1}^n z_i + \sum_{i=1}^n x_i}{n}$$

$$\begin{aligned}
 &= -\frac{1}{n}\widehat{a}_1 \sum_{i=1}^n z_i + \frac{1}{n} \sum_{i=1}^n x_i \\
 &= -\widehat{a}_1 \mu_Z + \mu_X
 \end{aligned} \tag{6.2.14}$$

und aus den Gleichungen (6.2.13) und (6.2.14) resultiert die Gleichung (6.2.7) für den Koeffizienten \widehat{a}_1 :

$$\begin{aligned}
 \sum_{i=1}^n x_i z_i &= \left(\frac{1}{n} \sum_{i=1}^n x_i - \widehat{a}_1 \frac{1}{n} \sum_{i=1}^n z_i \right) \cdot \sum_{i=1}^n z_i + \widehat{a}_1 \sum_{i=1}^n z_i^2 \\
 &= \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n z_i + \widehat{a}_1 \left(\sum_{i=1}^n z_i^2 - \frac{1}{n} \left(\sum_{i=1}^n z_i \right)^2 \right) \\
 \widehat{a}_1 &= \frac{\sum_{i=1}^n x_i z_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n z_i}{\sum_{i=1}^n z_i^2 - \frac{1}{n} \left(\sum_{i=1}^n z_i \right)^2} = \frac{(n-1) \sigma_{XZ}}{(n-1) \sigma_Z^2}
 \end{aligned}$$

In analoger Weise löst man das zweite Minimierungsproblem (6.2.6) für die Koeffizienten \widehat{b}_0 und \widehat{b}_1 . Die Bestimmung der partiellen Korrelation zwischen X und Y nach Elimination des linearen Anteils von Z erfolgt durch die Korrelation der Residuen U_{XZ} und U_{YZ}

$$r_{XY|Z} = \frac{COV(U_{XZ}, U_{YZ})}{\sqrt{VAR(U_{XZ})} \sqrt{VAR(U_{YZ})}}. \tag{6.2.15}$$

Die Gleichung (6.2.15) beschreibt die Korrelation X und Y , die verbleibt, wenn der lineare Einfluss von Z auf X und von Z auf Y berücksichtigt wird. Anders formuliert, der Einfluss von Z wurde herauspartialisiert. Unter Verwendung der Werte a_0, a_1, b_0, b_1 gilt für die Residuen

$$\begin{aligned}
 E(U_{XZ}) &= E(X - a_0 - a_1 Z) = 0, & VAR(U_{XZ}) &= \sigma_X^2 (1 - r_{XZ}^2) \\
 E(U_{YZ}) &= E(Y - b_0 - b_1 Z) = 0, & VAR(U_{YZ}) &= \sigma_Y^2 (1 - r_{YZ}^2)
 \end{aligned}$$

$$COV(U_{XZ}, U_{YZ}) = \sigma_X \sigma_Y (r_{XY} - r_{XZ} r_{YZ})$$

und damit ergibt sich die partielle Korrelation mit Gleichung (6.2.15) zu

$$r_{XY|Z} = \frac{r_{XY} - r_{XZ} r_{YZ}}{\sqrt{1 - r_{XZ}^2} \sqrt{1 - r_{YZ}^2}}.$$

Somit ist für normalverteilte Zufallsvariablen die bedingte Korrelation identisch mit der partiellen Korrelation, wie durch den Vergleich zu Gleichung (6.2.1) ersichtlich ist.

6.3 Partielle Korrelationen höherer Ordnung

Die partielle Korrelation bietet auch die Möglichkeit, dass der Einfluss von mehreren Variablen aus der Korrelation zwischen X und Y herauspartialsiert werden kann. Die sogenannte partielle Korrelation höherer Ordnung gibt also den linearen Zusammenhang zweier Variablen an, aus dem der lineare Einfluss weiterer Variablen eliminiert wurde. Wenn aus zwei Variablen nicht nur eine, sondern genau k von m Variablen herausgerechnet werden, so wird das eine partielle Korrelation k -ter Ordnung mit der Bezeichnung $r_{X_i X_j | X_1, \dots, X_k}$, $1 \leq i, j \leq m$, $k \in \{1, \dots, m\} \setminus \{i, j\}$ genannt. Demnach entspricht eine partielle Korrelation der Ordnung Null der gewöhnlichen (marginalen) Korrelation.[6] Für die partiellen Korrelationskoeffizienten gilt auch die Eigenschaft

$$-1 \leq r_{X_i X_j | X_1, \dots, X_k} \leq 1,$$

welche, wie bei der marginalen Korrelation, direkt aus der Cauchy-Schwarz Ungleichung folgt.[25]

Da vorerst alle Berechnungen des hier zu lösenden Problems auf den Exemplaren basieren, bestimmt man die partielle Korrelation zwischen je zwei Exemplaren X_i und X_j mit jeder möglichen Kombination der restlichen Menge von $m - 2$ Exemplare. Wenn alle partiellen Korrelation bis zur Ordnung k berechnet werden sollen, dann werden

$$\binom{m-2}{1} \cdot \binom{m-2}{2} \cdot \dots \cdot \binom{m-2}{k}$$

Berechnungen benötigt. Folglich kann der Rechenaufwand für höhere Ordnungen sehr hoch werden.

Für die Bestimmung der partiellen Korrelation gibt es drei Möglichkeiten, wobei dem Anwender überlassen ist, ob die Schätzung der Korrelation nach Pearson, Kendall oder Spearman erfolgen soll. Wie schon erwähnt wird hier der Pearsonsche Korrelationskoeffizient verwendet (siehe Begründung im Kapitel 6.1).

6.3.1 Berechnung der partiellen Korrelation mit linearer Regression

Bei der Berechnung der partiellen Korrelation k -ter Ordnung zwischen X_1 und X_2 mittels linearer Regression werden zunächst die folgenden Regressionsgleichungen bestimmt

$$\widehat{X}_1 = a_0 + a_1 X_3 + a_2 X_4 + \dots + a_{k-2} X_k,$$

$$\widehat{X}_2 = b_0 + b_1 X_3 + b_2 X_4 + \dots + b_{k-2} X_k.$$

Mit diesen Regressionsgleichungen werden \widehat{X}_1 und \widehat{X}_2 aufgrund der restlichen k Variablen X_3, \dots, X_k vorhergesagt. Die Varianzen dieser vorhergesagten Werte werden ausschließlich durch die Variablen X_3, \dots, X_k bestimmt. Aus der Subtraktion der vorhergesagten Werte von den tatsächlichen Werten resultieren die sogenannten bereinigten Werte X_1^* und X_2^* , deren Varianz von den k Variablen X_3, \dots, X_k unbeeinflusst sind.

$$X_1^* = X_1 - \widehat{X}_1 \tag{6.3.1}$$

$$X_2^* = X_2 - \widehat{X}_2 \tag{6.3.2}$$

Aus den Gleichungen (6.3.1) und (6.3.2), der bereinigten Variablen X_1^* und X_2^* , kann man die partielle Korrelation k -ter Ordnung mit

$$r_{X_1 X_2 | X_3, \dots, X_k} = \frac{COV(X_1^*, X_2^*)}{\sigma_{X_1^* X_2^*}}$$

berechnen. [6]

6.3.2 Rekursive Berechnung der partielle Korrelation

Bei der rekursiven Berechnung der partiellen Korrelation höherer Ordnung müssen zuvor sämtliche partiellen Korrelationen niedriger Ordnung bestimmt werden, wodurch der Rechenaufwand mit der Anzahl der Variablen sehr schnell ansteigt.

Die partielle Korrelation 1. Ordnung lässt sich durch die Korrelationen zwischen den beteiligten Variablen bestimmen. Für die Variablen X_1 und X_2 in Abhängigkeit von X_3 bedeutet das

$$r_{X_1 X_2 | X_3} = \frac{r_{X_1 X_2} - r_{X_1 X_3} \cdot r_{X_2 X_3}}{\sqrt{1 - r_{X_1 X_3}^2} \cdot \sqrt{1 - r_{X_2 X_3}^2}}.$$

In die partielle Korrelation 2. Ordnung gehen nur die partiellen Korrelationen 1. Ordnung ein. Diese lauten für die Variablen X_1 und X_2 in Abhängigkeit von X_3 und X_4 wie folgt

$$r_{X_1 X_2 | X_3 X_4} = \frac{r_{X_1 X_2 | X_3} - r_{X_1 X_4 | X_3} \cdot r_{X_2 X_4 | X_3}}{\sqrt{1 - r_{X_1 X_4 | X_3}^2} \cdot \sqrt{1 - r_{X_2 X_4 | X_3}^2}}.$$

Auf diesem Weg können alle partiellen Korrelationen für jede Ordnung bestimmt werden.[6]

6.3.3 Verwendung der inversen Korrelationsmatrix

Die Berechnungsmethode basierend auf der inversen Korrelationsmatrix ist der einfachste und schnellste Weg die partielle Korrelation k -ter Ordnung zu bestimmen. Es sei V die Menge aller Zufallsvariablen mit $|V| = m$. Die Berechnung der partiellen Korrelation $r_{X_i X_j | V \setminus \{X_i, X_j\}}$ basiert auf der Korrelationsmatrix der Menge $V \setminus \{X_i, X_j\}$. Unter der Voraussetzung, dass die Inverse dieser Korrelationsmatrix mit der Größe $(m - 2) \times (m - 2)$ existiert, kann die partielle Korrelation $r_{X_i X_j | V \setminus \{X_i, X_j\}}$ auf einfachem Weg berechnet werden.

Sei P die inverse Korrelationsmatrix von der Menge $V \setminus \{X_i, X_j\}$, dann gilt

$$r_{X_i X_j | V \setminus \{X_i, X_j\}} = -\frac{p_{ij}}{\sqrt{p_{ii} p_{jj}}}, \quad (6.3.3)$$

wobei p_{ij} die Matrixelemente von P bezeichnet.

Warum die partiellen Korrelationen auf dieser Art bestimmt werden können, ist auf den ersten Blick nicht ersichtlich. Für diesen Hintergrund wird der Begriff Konzentrationsmatrix benötigt. Die inverse Kovarianzmatrix Σ^{-1} wird als Konzentrationsmatrix bezeichnet. Seien c_{ij} die zugehörigen Elemente $\Sigma^{-1} = \{c_{ij}\}_{i,j=1}^m$. Bei Normalverteilung der Zufallsvariablen können die Matrixelemente c_{ij} folgendermaßen interpretiert werden: Die Einträge auf der Hauptdiagonalen der Konzentrationsmatrix Σ^{-1} sind die reziproken Varianzen der bedingten Verteilungen:

$$c_{ii} = \frac{1}{\text{VAR}(X_i | V \setminus X_i)}.$$

Das spezielle Diagonalelement c_{ii} ist damit der reziproke Wert der partiellen Varianz von

X_i unter Berücksichtigung der restlichen Menge $V \setminus X_i$.

Dagegen ist die Interpretation der Elemente abseits der Hauptdiagonalen nicht so offensichtlich. Wird die Konzentrationsmatrix auf gleiche Weise standardisiert, wie eine Kovarianzmatrix zur Korrelationsmatrix, dann beschreiben die Nichtdiagonalelemente c_{ij} die negative partielle Korrelation der beiden betreffenden Variablen X_i und X_j bedingt auf die übrige Menge $V \setminus \{X_i, X_j\}$

$$c_{ij} = -\sqrt{c_{ii}c_{jj}} \cdot r_{X_i X_j | V \setminus X_i X_j}. \quad (6.3.4)$$

Nach elementarer Umformung von Gleichung (6.3.4) erhält man eine Analogie zu Gleichung (6.3.3). Dass heißt, die Berechnung der partiellen Korrelation über die inverse Kovarianzmatrix liefert das gleiche Ergebnis, wie über die Methode basierend auf der inversen Korrelationsmatrix. [8]

Die Existenz der inversen Korrelationsmatrix für fMRT-Zeitreihen ist mit Wahrscheinlichkeit 1 gesichert, da, wie zuvor erwähnt, für die fMRT-Zeitreihen und damit auch für die Prototypen sowie die Exemplare im Allgemeinen angenommen werden kann, dass sie multivariat normalverteilt sind. Diese Existenzaussage begründet sich auf der Verteilung der Schätzung der Kovarianzmatrix. Für diese Verteilung wird eine spezielle multivariate Verteilung, die sogenannte Wishart-Verteilung, definiert.[27]

Die Wishart-Verteilung ist die multivariate Verallgemeinerung der χ^2 -Verteilung und kann als Summe von n unabhängigen standardnormalverteilten Zufallsvariablen definiert werden.

Definition 3. Wishart-Verteilung

Seien X_r , $r = 1, \dots, n$ unabhängig und multivariat normalverteilt $N_k(\mu_r, \Sigma)$. Dann nennt man die Verteilung der $k \times k$ Zufallsmatrix $W = \sum_{r=1}^n X_r X_r^T$ eine Wishart-Verteilung und W die Wishart-Matrix. Die Verteilung heißt zentral mit n Freiheitsgraden, wenn alle $\mu_r = 0$ sind und schreibt

$$W \sim W_k(\Sigma, n).$$

Andernfalls heißt die Verteilung nicht zentral mit

$$W \sim W_k(\Sigma, n, \lambda),$$

wobei $\lambda = (\mu_1, \dots, \mu_n)$.

Der folgende Satz trifft eine Aussage über den Verteilungstyp der geschätzten Kovarianzmatrix S_n .

Satz 1. *Seien X_1, \dots, X_n zufällig aus einer Grundgesamtheit gewählt, die $N_k(\mu, \Sigma)$ verteilt ist. Dann ist*

1. $\bar{x} \sim N_k(\mu, \Sigma/n)$
2. für $n \geq 2$ ist $S_n \sim W_k(\Sigma, n-1)$ und
3. \bar{x} und S_n sind unabhängig verteilt.

Die Wishart-Matrizen W sind für $n \geq k$ mit Wahrscheinlichkeit 1 positiv definit. Eine Matrix ist positiv definit, falls alle Eigenwerte größer Null sind. Damit ist Null kein Eigenwert von W und somit ist eine Wishart-Matrix invertierbar. Daraus folgt auch, dass die Wishart verteilte Kovarianzmatrix S_n invertierbar ist.[27]

Da sich die Korrelationen aus den Kovarianzen bestimmen und die partiellen Korrelationen auf der Basis der inversen Korrelationsmatrix resultieren, ist noch zu klären, welchen Verteilungstyp die inverse Korrelationsmatrix, dem zu folge die inverse Kovarianzmatrix hat. Dafür wird die inverse Wishart-Verteilung $W^{-1}(\Sigma^{-1}, n)$ eingeführt. Wenn $W \sim W_k(\Sigma, n)$ und $C = W^{-1}$, dann ist $C \sim W_k^{-1}(\Sigma^{-1}, n)$. [23]

6.3.4 Interpretation von Korrelationen

Die kausale Interpretation von Korrelationen kann leicht zu falschen Schlussfolgerungen führen. Im Abschnitt 5.2 wurde bereits erwähnt, dass im Allgemeinen gilt, dass $r_{X,Y} \neq r_{X,Y|Z}$. Es werden nachfolgend Fälle erläutert, in denen sich marginale und partielle Korrelationen unterscheiden.[3]

(1) Bedingte Unkorreliertheit (“Scheinkorrelationen”)

Eine Scheinkorrelation liegt dann vor, wenn zwei Zufallsvariablen X und Y beide von einer dritten Zufallsvariablen Z abhängen. Es könnte also sein, dass ein hoher Korrelationskoeffizient zwischen X und Y nur durch die Variable Z zustande kommt, genauer $r_{X,Y} \neq 0$ und $r_{X,Y|Z} = 0$. Des Weiteren besteht die Möglichkeit, dass die Variablen X

und Y nicht nur von einer Variablen, sondern von mehreren Variablen abhängig sind. Die Anzahl der Variablen die X und Y beeinflussen ist unbekannt.[3]

(2) Verdeckte Korrelationen

Im Gegensatz zur Scheinkorrelation kann der Fall eintreten, dass X und Y unkorreliert sind, obwohl die Variablen X und Y bedingt durchaus korreliert sind. Beispielsweise könnte eine Korrelation zwischen $X | Z$ und Y vorliegen und Unkorreliertheit zwischen X und Y . Es ist dann zweifelhaft die Variablen als nicht zusammenhängend zu charakterisieren, da sie tatsächlich in Untergruppen zusammenhängen. Formal bedeutet dass, wenn $r_{XY} = 0$ und $r_{XZ} \neq 0$ und $r_{YZ} \neq 0$, dann ergibt sich

$$r_{XY|Z} = \frac{-r_{XZ} \cdot r_{YZ}}{\sqrt{1 - r_{XZ}^2} \cdot \sqrt{1 - r_{YZ}^2}}$$

als partiellen Korrelationskoeffizienten.[3]

(3) Gekippte Effekte

Zwei Zufallsvariablen X und Y können positiv korreliert erscheinen bezüglich der marginalen Korrelation, obwohl die bedingte Korrelation $r_{XY|Z}$ einen negativen Zusammenhang (Korrelation) aufweist. Der Übergang von der marginalen Korrelation zur partiellen Korrelation kann eine Vorzeichenänderung zur Folge haben.[3]

(4) Supressor-Variablen

Es werden die Variablen X , Y und Z betrachtet. Wenn die partielle Korrelation $r_{XY|Z}$ stärker ist als die Korrelation r_{XY} , dann wird die Variable Z Supressor-Variable genannt. Dieses Phänomen tritt auf, wenn $r_{XZ} = 0$ oder $r_{YZ} = 0$ ist, da dann für die partielle Korrelation folgt, dass

$$r_{XY|Z} = \frac{r_{XZ}}{\sqrt{1 - r_{XZ}^2}}$$

oder

$$r_{XY|Z} = \frac{r_{XZ}}{\sqrt{1 - r_{YZ}^2}}$$

ist.[3]

6.3.5 Ein Scheinkorrelationsmaß

In der vorliegenden Arbeit sollen nur das Auftreten von Scheinkorrelationen für die fMRT-Zeitreihen genauer untersucht werden, also indirekten Verbindungen im Gehirn aufgedeckt werden. Ein Problem dabei ist ein Maß zu finden, ab wann der Unterschied zwischen der marginalen Korrelation $r_{X_i X_j}$ und der partiellen Korrelation k -ter Ordnung $r_{X_i X_j | X_1, \dots, X_k}$, $1 \leq i, j \leq m$, $k \in \{1, \dots, m\} \setminus \{i, j\}$, relevant ist. Bei der Berechnung der partiellen Korrelationen auf m Prototypen für die fMRT-Zeitreihen ist zu beobachten, dass der Wert $r_{X_i X_j | X_1, \dots, X_k}$, $1 \leq i, j \leq m$, $k \in \{1, \dots, m\} \setminus \{i, j\}$ kleiner wird, je höher die Ordnung k gewählt wird. In diesen Verläufen können keine Sprünge festgestellt werden, wodurch der Kurvenverlauf für ein Scheinkorrelationsmaß nicht geeignet ist.

Des Weiteren ist zu beachten, dass die Ordnung für die partiellen Korrelationen geeignet gewählt werden sollte. Wird bei einer Zeitreihenlänge von 168 die Ordnung 100 (entspricht 100 Voxeln) gewählt, so können keine sinnvollen Aussagen getroffen werden, da zuviel (fast vollständige) Information bei der partiellen Korrelation herausgerechnet wird. Im Allgemeinen sollte k nicht größer als $\frac{1}{5}$ der Freiheitsgrade der Zeitreihenlänge gewählt werden.[1]

Eine weitere Idee um ein Maß für Scheinkorrelationen zu entwickeln, ist die Verwendung der fMRT-Zeitreihen von zwei verschiedenen Gehirnen, da grundsätzlich angenommen wird, dass die Zeitreihen von zwei verschiedenen Gehirnen statistisch unkorreliert sind. In den zu untersuchenden fMRT-Zeitreihen wurden den Probanden keine Stimuli während des Experimentes präsentiert. Wenn allerdings den Probanden gleiche Stimuli präsentiert werden, dann könnten die fMRT-Zeitreihen zwar eventuell inhaltlich korreliert sein, aber auf Grund der nachfolgenden Erläuterung sollte Unkorreliertheit vorliegen. Diese Unkorreliertheit begründet sich auf die Phasenverschiebung und der unterschiedlichen Frequenz der fMRT-Zeitreihen, sowie dass das Untersuchungsgebiet der Gehirnnareale nicht exakt für zwei Gehirne lokalisiert werden kann.

Obwohl, die fMRT-Zeitreihen statistisch unkorreliert sein sollten, treten bei der Berechnung der Korrelationsmatrix signifikant von Null verschiedene Matrixelemente auf. Mit diesen Werten wird das Scheinkorrelationsmaß entwickelt, da es ein Hinweis auf Scheinkorrelationen sein kann. Durch die Bildung des Mittelwertes der von null verschiedenen Werte wird eine Schranke für die Scheinkorrelation und damit das Scheinkorrelationsmaß bestimmt. Genauer, wenn die Differenz zwischen r_{XY} und $r_{XY|Z}$ größer oder gleich dem Scheinkorrelationsmaß ist, dann kann angenommen werden, dass zwischen den Varia-

blen X und Y eine Scheinkorrelation vorliegt. Damit wird aus der Korrelationsmatrix $A_{Exemplare}$ (6.1.2) die neue Matrix $A_{Exemplare}^{partial}$ gebildet, indem bei vorhandener Scheinkorrelation diese neuen Werte (die partiellen Korrelationskoeffizienten) anstatt der originalen Korrelation von $A_{Exemplare}$ auftreten.

$$A_{Exemplare}^{partial} = \begin{pmatrix} 1 & 0.3058 & 0.2246 & 0.1975 & 0.0877 & 0.1962 & 0.3794 \\ 0.3058 & 1 & 0.6138 & 0.1147 & 0.2776 & 0.2592 & 0.2082 \\ 0.2246 & 0.6138 & 1 & 0.1014 & 0.4899 & 0.2401 & 0.1755 \\ 0.1975 & 0.1147 & 0.1014 & 1 & 0.3568 & 0.2580 & 0.4670 \\ 0.0877 & 0.2776 & 0.4899 & 0.3568 & 1 & 0.3623 & 0.0005 \\ 0.1962 & 0.2592 & 0.2401 & 0.2580 & 0.3623 & 1 & 0.3227 \\ 0.3794 & 0.2082 & 0.1755 & 0.4670 & 0.0005 & 0.3227 & 1 \end{pmatrix}. \quad (6.3.5)$$

Falls in der Matrix $A_{Exemplare}^{partial}$ negative Matrixelemente auftreten, wird eine Skalierung wie in Abschnitt 6.1 beschrieben, vorgenommen und damit die Voraussetzung für die Berechnung der Eigenvektor Zentralität erfüllt.

7 Partial Mutual Information

Bisher wurde in der Arbeit nur die Korrelation als Basis für die Berechnung der Eigenvektorzentralität in Betracht gezogen. Allerdings misst man mit diesem Ähnlichkeitsmaß, dem Korrelationskoeffizienten, nur die linearen Abhängigkeiten zwischen den fMRT-Zeitreihen. Da jedoch die Art der Abhängigkeit zwischen den fMRT-Zeitreihen unbekannt ist und damit auch nichtlinear sein kann, soll auch die Mutual Information als Ähnlichkeitsmaß in Betracht gezogen werden. In diesem Kapitel werden erst die theoretischen Grundlagen erläutert und anschließend wird auf die praktische Realisierung eingegangen.

7.1 Informationstheoretische Begriffe

Um die Theorie der Mutual Information beziehungsweise der Partial Mutual Information zu erläutern, werden zunächst einige grundlegende Begriffe aus der Informationstheorie eingeführt, welche für diesen Abschnitt aus den Büchern [29],[16] und [35] zusammengestellt wurden.

Den Grundstein dieser Theorie legte C.E. Shannon, er gilt auch als Begründer der Informationstheorie. Im Sinne der Nachrichtentechnik beschreibt die Informationstheorie die Übertragung und Codierung von Informationen. Durch die Beschreibung mittels Wahrscheinlichkeitsrechnung gelingt es, den Fluss der Information klar zu definieren.

Die Unsicherheit oder Entropie einer Zufallsvariable X , die mit Wahrscheinlichkeit p_i den Wert a_i annimmt ($1 \leq i \leq n$), hängt nur von den Wahrscheinlichkeiten p_1, \dots, p_n ab und wird mit $H_n(p_1, \dots, p_n)$ beziehungsweise $H(p_1, \dots, p_n)$ bezeichnet. Für $H_n(p_1, \dots, p_n)$ wird eine axiomatische Definition angegeben. Die Entropiefunktion $H_n := \{(p_1, \dots, p_n) \in [0, 1]^n \mid \sum_{i=1}^n p_i = 1\} \rightarrow \mathbb{R}$ soll folgende Eigenschaften aufweisen:

- (A1) $H_n(p_1, \dots, p_n)$ ist maximal genau dann, wenn, $p_1 = \dots = p_n = \frac{1}{n}$
- (A2) Für jede Permutation π von $1, \dots, n$ gilt:
 $H_n(p_1, \dots, p_n) = H_n(p_{\pi_1}, \dots, p_{\pi_n})$, dass heißt H_n ist eine sogenannte symmetrische Funktion.
- (A3) $H_n(p_1, \dots, p_n) \geq 0$ und es gilt $H_n(p_1, \dots, p_n) = 0 \iff p_i = 1$ für ein $i \in \{1, \dots, n\}$
- (A4) $H_{n+1}(p_1, \dots, p_n, 0) = H_n(p_1, \dots, p_n)$
- (A5) $H_n\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) \leq H_{n+1}\left(\frac{1}{n+1}, \frac{1}{n+1}, \dots, \frac{1}{n+1}\right)$ (Gleichheit existiert nicht, aber im Axiom ist die schwächere Bedingung besser)
- (A6) H_n ist stetig
- (A7) Für $n, m \in \mathbb{N}$ gilt:
 $H_{n+m}\left(\frac{1}{nm}, \frac{1}{nm}, \dots, \frac{1}{nm}\right) = H_m\left(\frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m}\right) + H_n\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$
- (A8) Seien $p = p_1 + \dots + p_m$, $q = q_1 + \dots + q_n$ mit $p_1, \dots, p_m \geq 0$, $q_1, \dots, q_n \geq 0$ und $p + q = 1$. Dann gilt
 $H(p_1, \dots, p_m, q_1, \dots, q_n) = H(p, q) + pH\left(\frac{p_1}{p}, \dots, \frac{p_m}{p}\right) + qH\left(\frac{q_1}{q}, \dots, \frac{q_n}{q}\right)$.

Satz 2. Es sei $H_n : \{(p_1, \dots, p_n) \in [0, 1]^n \mid \sum_{i=1}^n p_i = 1\} \rightarrow \mathbb{R}$ eine Funktion mit den Eigenschaften (A1) bis (A8). Dann gibt es eine positive reelle Zahl $\lambda > 0$, so dass für alle $(p_1, \dots, p_n) \in [0, 1]^n$ mit $\sum_{i=1}^n p_i = 1$ gilt

$$H_n(p_1, \dots, p_n) = -\lambda \sum_{k: p_k > 0} p_k \log p_k.$$

Definition 4. Entropiefunktion

Die Funktion $H_n : \{(p_1, \dots, p_n) \in [0, 1]^n \mid \sum_{i=1}^n p_i = 1\} \rightarrow \mathbb{R}$ mit

$H_n(p_1, \dots, p_n) = -\lambda \sum_{k: p_k > 0} p_k \log p_k$ heißt Entropiefunktion. H ist eine Verallgemeinerung von H_n auf der Menge aller diskreten Wahrscheinlichkeitsverteilungen.

$$H : \bigcup_{n \geq 1} \left\{ (p_1, \dots, p_n) \in [0, 1]^n \mid \sum_{i=1}^n p_i = 1 \right\} \rightarrow \mathbb{R}$$

Wegen $\lim_{x \rightarrow 0} x \log x = \lim_{x \rightarrow 0} \frac{\log x}{\frac{1}{x}} = \lim_{x \rightarrow 0} \frac{\frac{1}{x}}{\frac{-1}{x^2}} = \lim_{x \rightarrow 0} (-x) = -0 = 0$ kann man $0 \log 0 = 0$ definieren. Es sei $H(p_1, \dots, p_n) = -\sum_{i=1}^n p_i \log p_i$ eine Funktion mit den

Eigenschaften (A1) bis (A8). Damit kann die Entropie einer Zufallsvariablen X definiert werden.

Definition 5. Entropie und gemeinsame Entropie

Es sei X eine diskrete Zufallsvariable, die eine endliche Anzahl von Werten mit den Wahrscheinlichkeiten p_1, \dots, p_n annimmt. Dann heißt

$$H(X) := H(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log p_i$$

die Unsicherheit oder Entropie von X und wird auch Shannon Entropie genannt. Ist X von der Form $X = (X_1, \dots, X_m)$, dass heißt X ist ein Tupel bestehend aus m Zufallsvariablen, so schreibt man $H(X_1, \dots, X_m)$ statt $H((X_1, \dots, X_m))$ und nennt $H(X_1, \dots, X_m)$ die gemeinsame Entropie.

Die Entropie von X nimmt den Wert Null an, wenn die Zufallsvariable X konstant ist. Für konstante X gilt $p_k = 1$ für ein k und $p_i = 0$ für alle $i \neq k$. Aus $\log 1 = 0$ und $0 \log 0 = 0$ folgt $H(X) = 0$. Gilt umgekehrt $H(X) = 0$, so bedeutet das, dass $\sum_k p_k \log p_k = 0$. Folglich gilt für jedes k , dass $p_k \log p_k = 0$. Hieraus resultiert $p_k = 0$ oder $p_k = 1$. Da $\sum_k p_k = 1$, gilt $p_k = 1$ für genau ein k und $p_k = 0$ für alle anderen k . Demnach ist die Zufallsvariable X konstant.

Definition 6. Bedingte Entropie

Sei X eine Zufallsvariable mit Wertebereich $\{a_1, \dots, a_m\}$ und $A \in \Omega$ ein Ereignis mit $P(A) > 0$. Die Größe

$$H(X | A) := - \sum_{k=1}^m P(X = a_k | A) \log P(X = a_k | A)$$

heißt dann die bedingte Entropie von X unter A . Ist Y eine weitere Zufallsvariable mit den Werten $\{b_1, \dots, b_n\}$, so heißt

$$H(X | Y) := \sum_j H(X | Y = b_j) P(Y = b_j)$$

bedingte Entropie von X unter Y .

Die bedingte Entropie hat die folgenden Eigenschaften. Zum einen gilt $H(X | X) = 0$, da

$$\begin{aligned} H(X | X) &= \sum_j H(X | X = a_j) P(X = a_j) \\ &= - \sum_j \sum_k P(X = a_k | X = a_j) \log P(X = a_k | X = a_j) P(X = a_j) \\ &= - \sum_k P(X = a_k | X = a_k) \log P(X = a_k | X = a_k) P(X = a_k) = 0. \end{aligned}$$

Falls X und Y unabhängig sind dann gilt $H(X | Y) = H(X)$, weil

$$\begin{aligned} H(X | Y) &= \sum_j H(X | Y = b_j) P(Y = b_j) \\ &= - \sum_j \sum_k P(X = a_k | Y = b_j) \log P(X = a_k | Y = b_j) P(Y = b_j) \\ &= - \sum_j \sum_k P(X = a_k) \log P(X = a_k) P(Y = b_j) \\ &= - \sum_k P(X = a_k) \log P(X = a_k) \sum_j P(Y = b_j) = H(X). \end{aligned}$$

Weiterhin gilt für die bedingte Entropie von X unter Y , dass $H(X | Y) = 0$ genau dann, wenn $X = g(Y)$ für eine Funktion g .

Für je zwei Zufallsvariablen X und Y gilt $H(X, Y) = H(Y) + H(X | Y)$. Damit folgt, dass $H(X | Y) \leq H(X)$. Letzteres ist mit Gleichheit erfüllt genau dann, wenn X und Y unabhängig sind. Da

$$H(X | Y) = H(X, Y) - H(Y),$$

sowie $H(X, Y) \leq H(X) + H(Y)$ mit Gleichheit erfüllt ist genau dann, wenn für X und Y die Unabhängigkeit gegeben ist.

Definition 7. Information

Sei $E \in \Omega$ ein Ereignis mit $P(E) > 0$. Die Zahl

$$I(E) := -\log P(E)$$

heißt Information von E .

Es gilt $I(E) \geq 0$, sowie $I(E) = 0$ genau dann, wenn $P(E) = 1$. Je kleiner $P(E)$ ist, desto größer ist $I(E)$. Damit ist die Shannon-Entropie der Erwartungswert der Information.

Definition 8. Mutual Information

$I(X, Y)$ heißt die Information, die Y über X liefert, und ist definiert als

$$I(X, Y) = H(X) - H(X | Y).$$

$I(X, Y)$ bezeichnet man auch als wechselseitige Information, Transinformation oder Mutual Information.

Es ist somit $I(X, X) = H(X) - H(X | X)$.

Definition 9. Mittlere Information

Sei X eine Zufallsvariable mit endlich vielen Werten a_1, \dots, a_n . Dann gilt

$$\begin{aligned} H(X) &= - \sum_k P(X = a_k) \log_2 P(X = a_k) \\ &= \sum_k P(X = a_k) I(\{X = a_k\}). \end{aligned}$$

Dies entspricht der mittleren Information der Zufallsvariablen $\{X = a_k\} := I(X, X)$.

Des Weiteren gilt die Symmetrieeigenschaft $I(X, Y) = I(Y, X)$, welche folgendermaßen hergeleitet werden kann:

$$I(X, Y) = H(X) - H(X | Y)$$

$$= H(X) - [H(X, Y) - H(Y)],$$

aus $H(X, Y) = H(X | Y) + H(Y)$ und $H(X, Y) = H(Y | X) + H(X)$ folgt $-H(Y | X) = -H(X, Y) + H(X)$. Somit ergibt sich

$$I(X, Y) = -H(Y | X) + H(Y).$$

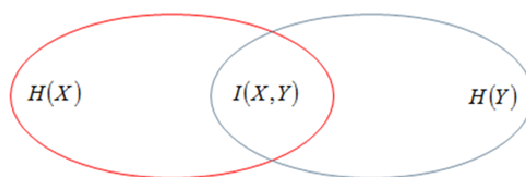


Abb. 7.2.1: Grafische Darstellung der Mutual Information für die Zufallsvariablen X und Y

Damit folgt die Symmetrieeigenschaft

$$I(X, Y) = I(Y, X).$$

7.2 Mutual Information

Aus den vorigen Abschnitt lässt sich zusammenfassen, dass die Mutual Information ein Maß für nichtlineare Abhängigkeiten zwischen 2 Zufallsvariablen ist. Die so eingeführte Mutual Information basiert auf der Shannon Entropie. Sie kann für zwei Zufallsvariablen X und Y auch als

$$I(X, Y) = H(X) + H(Y) - H(X, Y) \quad (7.2.1)$$

dargestellt werden.

Anschaulich kann $I(X, Y)$ in Gleichung (7.2.1) durch das Prinzip der Inklusion und Exklusion dargestellt werden (siehe Abbildung 7.2.1). Eine weitere wichtige Eigenschaft ist die Beschränktheit der Mutual Information, genauer $0 \leq I(X, Y) \leq \min\{H(X), H(Y)\}$. Es gilt $I(X, Y) = 0$ genau dann, wenn X und Y unabhängig sind. Denn es gilt $I(X, Y) = H(X) - H(X|Y)$ und $H(X) = H(X|Y)$ für unabhängige X und Y . Wenn Y eine Funktion von X ist gilt $I(X, Y) = H(Y)$. [10]

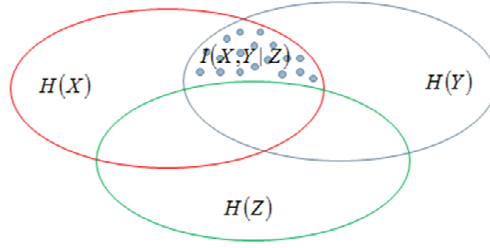


Abb. 7.3.1: Grafische Darstellung der Partial Mutual Information für die Zufallsvariablen X und Y unter der Bedingung Z

7.3 Partial Mutual Information

Analog zur partiellen Korrelation im Kapitel 6 wird in diesem Kapitel die Partial Mutual Information, welche die Information zwischen X und Y unter der Bedingung einer weiteren diskreten Zufallsvariablen Z bestimmt, eingeführt.

In der grafischen Darstellung 7.3.1 ist ersichtlich, dass nur der Teil von $I(X, Y)$ gesucht wird, welcher nicht in Z liegt. Die Partial Mutual Information ist mit den entsprechenden Entropiefunktionen wie folgt definiert

$$I(X, Y|Z) = H(X, Z) + H(Y, Z) - H(Z) - H(X, Y, Z).$$

Mit den gemeinsamen Wahrscheinlichkeiten $\{p_{x_i y_j z_k}\}_{i,j,k=1}^n$ von X , Y und Z und den entsprechenden Randwahrscheinlichkeiten $p_{x_i \cdot z_k} = \sum_j p_{x_i y_j z_k}$, $p_{\cdot y_j z_k} = \sum_i p_{x_i y_j z_k}$ und $p_{\cdot \cdot z_k} = \sum_{ij} p_{x_i y_j z_k}$ ergibt sich

$$I(X, Y|Z) = \sum_{ijk} p_{x_i y_j z_k} \ln \left(\frac{p_{x_i y_j z_k} p_{\cdot \cdot z_k}}{p_{x_i \cdot z_k} p_{\cdot y_j z_k}} \right). \quad (7.3.1)$$

Die Partial Mutual Information ist ebenso wie die Mutual Information symmetrisch, $I(X, Y|Z) = I(Y, X|Z)$. Ebenso gilt $0 \leq I(X, Y|Z)$, wobei $I(X, Y|Z) = 0$ für genau drei Fälle:

1. $I(X, Y|Z) = 0$ genau dann, wenn X und Y unabhängig sind unter der Bedingung Z , dass heißt, wenn für alle X, Y, Z gilt: $\left(\frac{p_{x_i y_j z_k}}{p_{\cdot \cdot z_k}} \right) = \left(\frac{p_{x_i \cdot z_k}}{p_{\cdot \cdot z_k}} \right) \left(\frac{p_{\cdot y_j z_k}}{p_{\cdot \cdot z_k}} \right)$.

2. $I(X, Y | Z) = 0$ wenn X unabhängig von Z ist, Y jedoch abhängig von Z .
3. $I(X, Y | Z) = 0$, wenn X abhängig von Z ist, Y jedoch unabhängig von Z .

Sind X, Y und Z kontinuierliche Zufallsvariablen mit der Wahrscheinlichkeitsverteilung $p(x, y, z)$ und den Randdichten $p(x, \cdot, z), p(\cdot, y, z)$ und $p(\cdot, \cdot, z)$, dann ergibt sich die Partial Mutual Information zu

$$I(X, Y | Z) = \int p(x, y, z) \ln \frac{p(x, y, z) p(\cdot, \cdot, z)}{p(x, \cdot, z) p(\cdot, y, z)} dx dy dz. \quad (7.3.2)$$

Im kontinuierlichen Fall heißt

$$h(X) = - \int p(x) \ln(p(x)) dx$$

Differential Entropie für jede kontinuierliche Zufallsvariable X mit Dichte $p(x)$. Somit kann die Gleichung (7.3.2) unter Zuhilfenahme der Differential Entropie formuliert werden

$$I(X, Y | Z) = h(X, Z) + h(Y, Z) - h(Z) - h(X, Y, Z). \quad (7.3.3)$$

Die kontinuierliche Partial Mutual Information ist invariant gegenüber streng monotonen Transformationen, dass heißt, sie ist robust gegenüber möglichen nichtlinearen Verformungen von Zeitreihen. Das ist ein Vorteil gegenüber der Korrelation, welche diese Eigenschaft nicht erfüllt. Wenn die Zufallsvariable Z vernachlässigt werden kann, das heißt, wenn X und Y unabhängig von Z sind, dann ist die Partial Mutual Information äquivalent zur Mutual Information. Das kann folgendermaßen hergeleitet werden: ausgehend von Gleichung (7.3.3) gilt für von Z unabhängige X und Y , dass

$$I(X, Y | Z) = h(X) + h(Z) + h(Y) + h(Z) - h(Z) - (h(X, Y) - h(Z))$$

ist und somit folgt

$$I(X, Y) = h(X) + h(Y) - h(X, Y).$$

Bildlich veranschaulicht kann, man sich das so vorstellen, dass in der Abbildung (7.3.1) $H(Z)$ entfernt wird und es verbleibt $I(X, Y)$. [10]

7.3.1 Schätzung der Partial Mutual Information

Die Mutual Information sowie die Partial Mutual Information können nur unter der Voraussetzung der gegebenen Wahrscheinlichkeitsdichten der Zeitreihen bestimmt werden, da die Berechnung der Entropiefunktion nach Shannon auf diesen Wahrscheinlichkeiten basiert. Das Problem ist jedoch, dass für die vorliegenden fMRT-Zeitreihen diese benötigten Wahrscheinlichkeitsdichten nicht vorliegen. Ein Ausweg bietet die Schätzung der Entropiefunktionen.

7.3.1.1 Entropieschätzung

Es wird angenommen, dass die Datenpunkte (hier Zeitreihen) im \mathbb{R}^n liegen. Ausgangspunkt ist eine Überdeckung des \mathbb{R}^n mit identisch großen n -dimensionalen Würfeln (Boxen). [12]

Ausgangspunkt der Schätzung ist die Betrachtung von $M \gg 1$ Boxen und $N \gg 1$ zufälligen unabhängig verteilten Punkten in den Boxen, wobei jede Box das Gewicht p_i , $i = 1, \dots, M$ mit $\sum_{i=1}^M p_i = 1$ hat. Jede dieser Boxen enthält eine zufällige Anzahl von Punkten n_i mit $E(n_i) = p_i N$. Die Anzahl der Punkte in einer bestimmten Box genügt einer Binomialverteilung.

$$P(n_i, p_i, N) = \binom{N}{n_i} p_i^{n_i} (1 - p_i)^{N - n_i}$$

Für alle p_i wird angenommen, dass $p_i \ll 1$ für alle i gilt. Damit sind die n_i annähernd Poisson verteilt

$$P(n_i, z_i) = \frac{z_i^{n_i}}{n_i!} e^{-z_i}$$

mit dem Parameter $z_i \equiv E(n_i) = p_i N$. Das Ziel ist die Schätzung der Entropie

$$H = - \sum_{i=1}^M p_i \ln p_i = \ln N - \frac{1}{N} \sum_{i=1}^M z_i \ln z_i$$

der beobachteten Zahlen $\{n_i\}$. Wenn man einen naiven Schätzer \widehat{H}_{naiv} mit $z_i = n_i$ ansetzt, dann erhält man

$$\widehat{H}_{naiv} = \ln N - \frac{1}{n} \sum_{i=1}^M n_i \ln n_i$$

eine Unterschätzung, dass heißt $E(\widehat{H}_{naiv}) - H < 0$. Für große N und M ist die Einbringung von $z_i \ln z_i$ zur Entropie statistisch unabhängig und kann damit unabhängig vom Schätzer, welches allein eine Funktion von n_i ist, berechnet werden. Diese Einbringung wird mit $\phi(n_i)$ bezeichnet und stellt eine noch unbekannte Ansatzfunktion dar.

$$z_i \ln z_i \approx \widehat{n_i \ln n_i} = n_i \phi(n_i) \quad (7.3.4)$$

Der Erwartungswert dafür ist

$$E(\widehat{z_i \ln z_i}) = \sum_{n_i=1}^M n_i \phi(n_i) P(n_i, z_i). \quad (7.3.5)$$

Das bedeutet, man erhält unter der Annahme, dass für ein spezielles Ergebnis $n_i = 0$ gilt, dass es keine Information gibt. Wenn es ein spezielles $n_i = 0$ gibt, dann liegt der Grund dafür entweder in statistischen Schwankungen oder das für eine Box i $p_i = 0$ gilt. Es resultiert der Entropieschätzer

$$\widehat{H}_\phi = \ln N - \frac{M}{N} \overline{n \phi(n)}, \quad (7.3.6)$$

wobei $\overline{n \phi(n)} = \frac{1}{n} \sum_{i=1}^n n_i \phi(n_i)$ der Mittelwert über alle Boxen ist.

Eine Vereinfachung bringt die Betrachtung der allgemeineren Rényi Entropie statt der Shannon Entropie:

$$\begin{aligned} H(q) &= \frac{1}{1-q} \ln \sum_{i=1}^M p_i^q \\ &= \frac{1}{1-q} \left[\ln \sum_{i=1}^M z_i^q - q \ln N \right]. \end{aligned}$$

Für $q \rightarrow 1$ resultiert wieder die Shannon Entropie. In den Gleichungen (7.3.4), (7.3.5) und (7.3.6) ersetzt man $\widehat{z_i^q} = n_i \phi(n_i, q)$ mit $\phi(n) = \frac{d\phi(n, q)}{dq} \big|_{q=1}$,

$$E(\widehat{z_i^q}) = \sum_{n_i} n_i \phi(n_i, q) P(n_i, z_i).$$

Für $q \geq 2$ ist der fehlerfreie Schätzer gegeben durch

$$\widehat{z^q} = \frac{n!}{(n-q)!}$$

und aus dem Erwartungswert ergibt sich

$$\sum_{n=q}^{\infty} \frac{n!}{(n-q)!} P(n, z) = z^q. \quad (7.3.7)$$

Eine gute Strategie ist, zuerst die Gleichung (7.3.7) auf der linken Seite für beliebige q zu verallgemeinern und danach den Unterschied zu z^q näher zu analysieren. Für eine spätere analytische Verallgemeinerung startet man mit negativen q und geht zu positive q über. Es wird definiert

$$\begin{aligned} A(-q, z) &= \sum_{n=1}^{\infty} \frac{\Gamma(n+1)}{\Gamma(n+1+q)} \frac{z^n}{n!} e^{-z} \\ &\equiv E\left(\frac{\Gamma(n+1)}{\Gamma(n+1+q)}\right). \end{aligned} \quad (7.3.8)$$

Es gilt $\frac{\Gamma(n+1)}{\Gamma(n+1+q)} = \frac{B(n+1, q)}{\Gamma(q)}$ und für die Betafunktion B verwendet man die Integraldarstellung

$$B(n+1, q) = \int_0^1 (1-t) t^{q-1} dt. \quad (7.3.9)$$

Da das Integral in (7.3.9) und die Summe in (7.3.8) konvergiert, kann es ausgetauscht werden. Die Summe wird dann zu

$$\begin{aligned} A(-q, z) &= \frac{1}{\Gamma(q)} \int_0^1 t (e^{-tz} - e^{-z}) dt \\ &= \frac{z^{-q}}{\Gamma(q)} \int_0^z x^{q-1} e^{-x} - \frac{e^{-x}}{\Gamma(1+q)} dt. \end{aligned} \quad (7.3.10)$$

Der letzte Term entsteht daher, dass die Summe über alle n nur von 1 bis ∞ beinhaltet. Durch Aufspaltung des Integrals $\int_0^z = \int_0^\infty - \int_z^\infty$ kann der erste Term in (7.3.10) als Gammafunktion und die zweite als unvollständige Gammafunktion ausgedrückt werden.

$$A(-q, z) = z^{-q} - \frac{z^{-q}}{\Gamma(q)} \Gamma(q, z) - \frac{e^{-z}}{\Gamma(1+q)}$$

Hier kann schließlich die Verallgemeinerung für positive q erfolgen. Nachfolgend wird folgende Rekursionsbeziehung der Gammafunktion verwendet

$$\Gamma(a, z) = \frac{1}{a} \Gamma(1+a, z) - \frac{e^{-z} z^a}{a}$$

um

$$E\left(\frac{\Gamma(n+1)}{\Gamma(n+1-q)}\right) = z^q - \frac{z^q}{\Gamma(1-q)}\Gamma(1-q, z) \quad (7.3.11)$$

zu erhalten. Für die Shannon-Entropie nimmt man die erste Ableitung von (7.3.11) bezüglich q , setzt $q = 1$ und es resultiert

$$E(n\psi(n)) = z \ln z + zE_1(z).$$

Hier ist $\psi(x) = \frac{d \ln \Gamma(x)}{dx}$ die Digammafunktion und $E_1(x) = \Gamma(0, x) = \int_1^\infty \frac{e^{-xt}}{t} dt$ ein Exponential Integral ist. Für große z gilt $zE_1(x) \approx e^{-z}$. Mit der Digammafunktion resultiert der für uns interessante Entropieschätzer \widehat{H}_ψ ,

$$\widehat{H}_\psi = \ln N - \frac{1}{N} \sum_{i=1}^M n_i \psi(n_i).$$

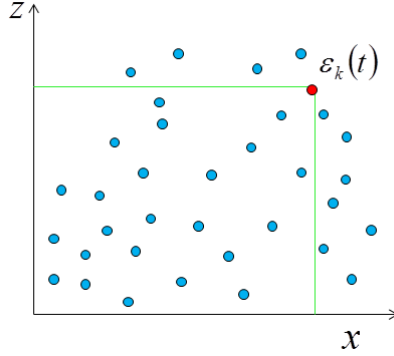
Besonders hervorzuheben ist die Approximation des natürlichen Logarithmus mit der Digammafunktion, $\psi(x) \approx \ln x$ [12]. Die Mutual Information von zwei Zufallsvariablen X und Y kann nun mit der Digammafunktion bestimmt werden. Dabei steht die Digammafunktion in Verbindung zur harmonische Zahl

$$h_n = \psi(n+1) - \psi(1). \quad (7.3.12)$$

[15] Diese Beziehung ist der wesentliche Punkt, für den Hintergrund der nachfolgenden Schätzmethode für die Mutual Information, die auf dem k -nächsten Nachbar basiert.

7.3.1.2 Entropieschätzung basierend auf den k -nächsten Nachbar

Die Partial Mutual Information sowie die Mutual Information wird mit einer Methode basierend auf k -nächsten Nachbar geschätzt, welche von Kraskov, Stögbauer und Grassberger in [17] vorgeschlagen wurde. Die Theorie dieses Abschnitts für diese Methode wurde aus den Veröffentlichungen [10] und [17] entnommen. Ausgangspunkt für die Schätzung sind drei Zeitreihen $\{x_t\}$, $\{y_t\}$ und $\{z_t\}$, welche endliche Realisierungen von stationären ergodischen Prozessen $\{X_t\}$, $\{Y_t\}$ und $\{Z_t\}$ sind, die einen Raum aufspannen. Für jeden Vektor $v_t = (x_t, y_t, z_t)$, $t = 1, 2, \dots, T$, und für jedes feste k mit $1 \leq k \leq T$ wird die Distanz $\varepsilon_k(t)$ zum k -nächsten Nachbar bestimmt. Demnach sind in $\{v_{t^*}\}$ mit $t^* = 1, \dots, T$, $t^* \neq t$ genau $k-1$ Punkte mit der Distanz streng kleiner als $\varepsilon_k(t)$ und $N - k - 1$ Punkte


 Abb. 7.3.2: grafische Darstellung von N_{xz}

streng größer als $\varepsilon_k(t)$. Die Berechnung der Distanz erfolgt mit der Maximumnorm. Die grundlegende Idee ist, $H(X)$ durch den mittleren Abstand zum k -nächsten Nachbar gemittelt über alle x_i zu bestimmen. Anschließend werden Randvektoren $w_t^{xz} = (x_t, z_t)$, $t = 1, 2, \dots, T$ festgelegt, die eine Ebene aufspannen. In der Ebene zählt man für jedes t die Anzahl der Punkte in $\{w_{t^*}^{xz}\}$, welche eine Distanz streng kleiner als $\varepsilon_k(t)$ haben (Abbildung), bezeichnet mit N_{xz} .

$$N_{xz}(t) = \# \{t^* \neq t : \|w_{t^*}^{xz} - w_t^{xz}\| < \varepsilon_k(t)\}$$

Zu beachten ist, dass $k \leq N_{xz} + 1$ gilt. Analog verhält sich die Bestimmung von $N_{yz}(t)$ und $N_z(t)$ mit den Randvektoren $w_t^{yz} = (y_t, z_t)$ und $w_t^{zz} = (z_t, z_t)$:

$$N_{yz}(t) = \# \{t^* \neq t : \|w_{t^*}^{yz} - w_t^{yz}\| < \varepsilon_k(t)\}$$

$$N_{zz}(t) = \# \{t^* \neq t : \|w_{t^*}^{zz} - w_t^{zz}\| < \varepsilon_k(t)\}$$

Die Schätzung $I(X, Y|Z)$ ergibt sich nun wie folgt:

$$\hat{I}(X, Y|Z) = \frac{1}{T} \sum_{t=1}^T (h_{N_{xz}(t)} + h_{N_{yz}(t)} - h_{N_z(t)}) - h_{k-1}, \quad (7.3.13)$$

wobei h_N die n -te negative harmonische Zahl mit $h_N = -\sum_{n=1}^N \frac{1}{n}$ ist. Wenn die Bedingung Z vernachlässigt wird, das heißt X und Y unabhängig von Z sind, dann folgt $N_z = T - 1$, $N_{xz} = N_x$ und $N_{yz} = N_y$. Dies führt zur Mutual Information $I(X, Y)$ und ergibt für die

Schätzung

$$\hat{I}(X, Y) = \frac{1}{T} \sum_{t=1}^T (h_{N_x(t)} + h_{N_y(t)}) - h_{T-1} - h_{k-1}.$$

Hinweise zur Wahl von k

Bei der Entropieschätzung basierend auf den k -nächsten Nachbar sollte k sorgfältig gewählt werden. Untersuchungen ergaben, dass eine Wahl entsprechend 2% – 3% der Zeitreihenlänge stabile Ergebnisse für die Schätzung der Mutual Information liefert.[9] Bei ungeeigneten Parameter k können negative Werte auftreten. Dieser Fall ist zu vermeiden, da nach der Definition der Mutual Information $I(X, Y|Z) \geq 0$ gilt. Somit kann eine Schätzung basierend auf ungeeigneten k nicht für weitere Untersuchungen verwendet werden. Eine weitere Folgerung ist, dass die Differenz zwischen $I(X, Y|Z)$ und $\hat{I}(X, Y|Z)$ bei ungeeignetem k zu groß werden kann. Was als Nachteil zu betrachten ist, da dann die Schätzung der Mutual Information zu stark von der “tatsächlichen” Mutual Information abweichen würde.

Ein weiteres Problem besteht darin, dass die Schätzung der Mutual Information in Gleichung (7.3.13) auf den harmonischen Zahlen basiert und diese für den Fall $N = 0$ nicht definiert sind. Dieser Fall tritt bei der Berechnung der Partial Mutual Information nach Gleichung (7.3.13) für $k = 1$ ein. Für diesen Fall muss die harmonische Zahl h_0 besonders behandelt werden. Dafür werden folgende Überlegungen angestellt:

1. Die harmonische Zahl steht in Beziehung zur Digammafunktion, siehe Gleichung (7.3.12). Dann folgt mit $n = 0$, dass $h_0 = \psi(1) - \psi(1) = 0$.
2. Basierend auf der rekursiven Darstellung der negativen harmonischen Zahlen

$$h_k = h_{k-1} - \frac{1}{k}$$

gilt für $k = 1$

$$h_1 = h_0 - \frac{1}{1}$$

Somit ist $-1 = h_0 - 1$. Damit gilt wieder $h_0 = 0$.

7.3.2 Eine Ersatzmethode für den Signifikanztest

Die Berechnung der Mutual Information $I(X, Y)$ und der Partial Mutual Information $I(X, Y|Z)$ erfolgt lediglich auf m Exemplaren der n fMRT-Zeitreihen. Die Exemplare resultierten aus der Clusterlösung mit Affinity Propagation basierend auf den Prototypen vom Neural Gas Algorithmus.

Auf die resultierenden Matrizen kann kein t-Test (Signifikanztest) angewendet werden, da die Werte $\hat{I}(X, Y)$ und $\hat{I}(X, Y | Z)$ nicht normalverteilt sind. Demzufolge sind in der Matrix Werte nahe Null enthalten, die eventuell vernachlässigbar sind. Im Gegensatz dazu beinhalten die Korrelationsmatrizen nach Anwendung eines t-Tests nur noch signifikant von Null verschiedene Werte. Bei der Schätzung der Mutual Information geht aber keine explizite Verteilungsinformation ein, wodurch die Anwendung eines t-Test nicht möglich ist und ein Ersatz dafür zu entwickeln ist. Es ist also erforderlich ein Maß zu konstruieren, so dass nur die wesentlichen Werte ungleich Null in den Matrizen verbleiben. Es wird nun eine Methode vorgestellt, welche an sich kein Signifikanztest ist, sondern einen Ersatz für den Signifikanztest darstellen soll.

Die Idee besteht darin, dass die Mutual Information für zwei unterschiedliche Zeitreihen x_i und x_j gleicher Länge den Wert $I(x_i, x_j) = 0$ annehmen soll, da diese, wie vorher beschrieben, als unabhängig angenommen werden. Tatsächlich ist der resultierende Wert aus der Schätzung der Mutual Information nahe Null. Zur praktischen Konstruktion eines Maßes nimmt man sich eine fMRT-Zeitreihe y aus einem neuen Datensatz, welche von dem vorliegenden Datensatz verschieden ist. Diese Zeitreihe y und die Zeitreihen $\{x_1, \dots, x_n\}$ des bekannten Datensatzes haben unterschiedliche Frequenzen und Phasenverschiebungen. Der Mittelwert der Werte $I(x_i, y)$, $i = 1, \dots, n$, die von Null abweichen, bildet das gewünschte Maß. Dieses Maß soll den Ersatz für das Signifikanzmaß darstellen. Damit können alle Werte, die dem Betrage nach kleiner als das Maß sind, auf Null gesetzt werden.

Auf Grund dieser Überlegung kann nun eine Matrix analog zur Korrelationsmatrix angegeben werden, welche hier Mutual Information Matrix genannt wird. Diese Matrix enthält die Werte für die Mutual Information zwischen je zwei Exemplaren und wird mit $MI_{Exemplare}$ bezeichnet.

$$MI_{Exemplare} = \begin{pmatrix} 0 & 0,4289 & 0,3839 & 0,3801 & 0,3908 & 0,3641 & 0,2716 \\ 0,4289 & 0 & 0,6565 & 0,5318 & 0,3682 & 0,3015 & 0,4022 \\ 0,3839 & 0,6565 & 0 & 0,4368 & 0,5134 & 0,4039 & 0,4502 \\ 0,3801 & 0,5318 & 0,4368 & 0 & 0,4969 & 0,3697 & 0,3426 \\ 0,3908 & 0,3682 & 0,5134 & 0,4969 & 0 & 0,5018 & 0,3619 \\ 0,3641 & 0,3015 & 0,4039 & 0,3697 & 0,5018 & 0 & 0,3304 \\ 0,2716 & 0,4022 & 0,4502 & 0,3426 & 0,3619 & 0,3304 & 0 \end{pmatrix} \quad (7.3.14)$$

Ebenso wie die Matrix $A_{Exemplare}$ enthält auch die Matrix $MI_{Exemplare}$ relativ kleine Werte. Dieses Resultat begründet sich darauf, dass jeweils ähnliche fMRT-Zeitreihen ein Cluster bilden (siehe Kapitel 6.1.1).

7.3.3 Pseudo Mutual Information

In Analogie zur Scheinkorrelation wird hier der Begriff Pseudo Mutual Information eingeführt. Eine Pseudo Mutual Information liegt dann vor, wenn zwei Zufallsvariablen X und Y in Abhängigkeit zu einer dritten Zufallsvariablen Z stehen. Dass bedeutet, dass ein hoher Wert $\hat{I}(X, Y)$ durch die Abhängigkeit zu Z verursacht werden könnte, genauer $\hat{I}(X, Y) \neq 0$ und $\hat{I}(X, Y | Z) = 0$. Auch hier besteht das Problem darin zu entscheiden, ab wann die Differenz zwischen $\hat{I}(X, Y)$ und $\hat{I}(X, Y | Z)$ als relevant zu betrachten ist. Es stellt sich also die Frage: Wann liegt eine Pseudo Mutual Information vor?

Dafür wird wie bei der Scheinkorrelation ein Maß entwickelt, welches angibt, ab wann die Differenz zwischen $\hat{I}(X, Y)$ und $\hat{I}(X, Y | Z)$ als relevant zu betrachten ist. Das dafür vorgestellte Verfahren ist eine Erweiterung der Methode für den Signifikanztest für die Mutual Information. Der Ausgangspunkt für diese Vorgehensweise ist die Überlegung, dass $\hat{I}(x_i, x_j | y) = 0$, wenn die fMRT-Zeitreihen $x_i, x_j \in \{x_1, \dots, x_n\}$ und die Zeitreihe y aus zwei verschiedenen Gehirnen stammen, wegen unterschiedlicher Frequenz und Phasenverschiebung. Dabei sind die Zeitreihen x_i, x_j aus dem vorliegenden Datensatz und die Zeitreihe y aus einem weiteren Datensatz. Für das gesuchte Maß bestimmt man $\hat{I}(x_i, x_j | y)$ zwischen jedem Paar $i, j \in \{1, \dots, n\}$ und anschließend den Mittelwert. Dass heißt, der Mittelwert über alle $\hat{I}(x_i, x_j | y), i, j \in \{1, \dots, n\}$ bildet das Maß für die Pseudo Mutual Information.

Des Weiteren kann eine Pseudo Mutual Information auch durch die Abhängigkeit von mehr als einer Variablen Z verursacht werden, wobei die Anzahl der Variablen, die X und Y beeinflussen, unbekannt ist.

7.3.4 Skalierungsmethode

Nach der Aussage des Satzes von Perron-Frobenius setzt die Bestimmung zentraler Knoten mit der Eigenvektorzentralität eine Ähnlichkeitsmatrix A mit ausschließlich positiven Einträgen $a_{ij}, \forall i, j$ voraus (siehe Kapitel 3.2). Die Mutual oder Partial Mutual Information kann man als Ähnlichkeitsmaß verwenden. Theoretisch gilt zwar, dass $I(X, Y) \geq 0$, sowie $I(X, Y | Z) \geq 0$, aber durch die Schätzung dieser Werte mit dem k nächsten Nachbar können $\hat{I}(X, Y)$ und $\hat{I}(X, Y | Z)$ auch negative Werte annehmen. Eine Skalierung mit der Addition von Eins, wie man bei den Korrelationsmatrizen vornimmt, reicht an dieser Stelle nicht aus, da $\hat{I}(X, Y)$ beziehungsweise $\hat{I}(X, Y | Z)$ Werte größer 1 annehmen können. Allerdings kann man, falls negative Werte vorliegen, durch die Minimumbildung aller negativen Werte, in den Matrizen für die Mutual und Partial Mutual Information und anschließender Addition des Minimumwertes die Ähnlichkeitsmatrix skalieren, dass heißt $\widetilde{a}_{ij} = a_{ij} - \min_{i,j} a_{ij}, \forall i, j$, mit $a_{ij} < 0$, da die Addition einer Konstanten die Mutual Information nicht ändert.

7.4 Partial Mutual Information höherer Ordnung

Bisher wurde bei der Partial Mutual Information nur der Einfluss einer weiteren Variablen Z auf die Variablen X und Y berücksichtigt. Die Partial Mutual Information höherer Ordnung bietet die Möglichkeit, die nichtlinearen Abhängigkeiten von mehreren Variablen bezüglich X und Y zu beachten. Wenn die Partial Mutual Information von zwei Variablen X_i und X_j , $1 \leq i, j \leq m$ genau in Abhängigkeit von k , $k \in \{1, \dots, m\} \setminus \{i, j\}$ Variablen bestimmt wird, dann spricht man von einer Partial Mutual Information k -ter Ordnung und man schreibt $I(X_i, X_j | X_{i_1}, \dots, X_{i_k})$. Folglich entspricht eine Partial Mutual Information der Ordnung Null der regulären Mutual Information.

Äquivalent zu den Korrelationen basieren alle Berechnungen auf den Exemplaren, wobei die Exemplare Zeitreihen sind und einer Verteilung unterliegen. Demnach werden zwischen je zwei Exemplaren X_i und X_j und jeder möglichen Kombination der restlichen Menge von $m - 2$ Exemplaren die Partial Mutual Information k -ter Ordnung bestimmt. Die Anzahl der Berechnungen bis zur Ordnung k beträgt wie bei den partiellen Korrelationen k -ter Ordnung

$$\binom{m-2}{1} \cdot \binom{m-2}{2} \cdot \dots \cdot \binom{m-2}{k}.$$

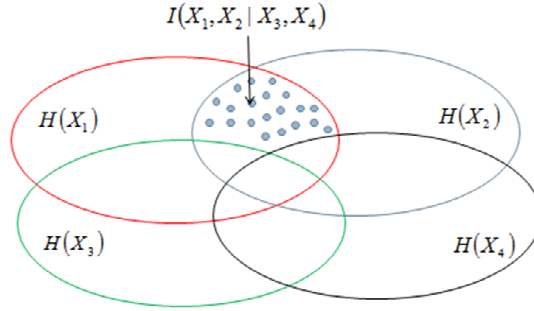


Figure 7.4.1: Grafische Darstellung der Partial Mutual Information 2-ter Ordnung

Es sei V wieder die Menge aller Zufallsvariablen (Exemplare) mit $|V| = m$ und $\tilde{V} = V \setminus \{X_i, X_j\}$. Bei der Partial Mutual Information k -ter Ordnung wird die Information zwischen X_i und X_j , $1 \leq i, j \leq m$, unter der Bedingung $k \in \{1, \dots, m\} \setminus \{i, j\}$ weiterer Zufallsvariablen $X_{i_1}, \dots, X_{i_k} \in \tilde{V}$ bestimmt. Von $I(X_i, X_j)$ ist nur der Bereich von Interesse, welcher nicht in $\{X_{i_1}, \dots, X_{i_k}\}$ liegt. Mit den zugehörigen Entropiefunktionen ist die Partial Mutual Information k -ter Ordnung wie folgt definiert

$$\begin{aligned} I(X_i, X_j | X_{i_1}, \dots, X_{i_k}) &= H(X_i, X_{i_1}, \dots, X_{i_k}) + H(X_j, X_{i_1}, \dots, X_{i_k}) \\ &\quad - H(X_{i_1}, \dots, X_{i_k}) - H(X_i, X_j, X_{i_1}, \dots, X_{i_k}). \end{aligned}$$

Für diskrete Zufallsvariablen mit den gemeinsamen Wahrscheinlichkeiten $\{p_{X_i X_j X_{i_1}, \dots, X_{i_k}}\}$ von $X_i, X_j, X_{i_1} \dots X_{i_k}$, sowie den entsprechenden Randwahrscheinlichkeiten

$$p_{X_i \cdot X_{i_1}, \dots, X_{i_k}} = \sum_{X_j} p_{X_i X_j X_{i_1}, \dots, X_{i_k}},$$

$$p_{\cdot X_j X_{i_1}, \dots, X_{i_k}} = \sum_{X_i} p_{X_i X_j X_{i_1}, \dots, X_{i_k}}$$

und

$$p_{\cdot \cdot X_{i_1}, \dots, X_{i_k}} = \sum_{X_i X_j} p_{X_i X_j X_{i_1}, \dots, X_{i_k}}$$

resultiert

$$I(X_i, X_j | X_{i_1}, \dots, X_{i_k}) = \sum_{X_i X_j X_{i_1} \dots X_{i_k}} p_{X_i X_j X_{i_1} \dots X_{i_k}} \cdot \ln \left(\frac{p_{X_i X_j X_{i_1} \dots X_{i_k}} p_{\cdot X_{i_1} \dots X_{i_k}}}{p_{X_i \cdot X_{i_1} \dots X_{i_k}} p_{\cdot X_j X_{i_1} \dots X_{i_k}}} \right). \quad (7.4.1)$$

Es gilt ebenfalls die Symmetrieeigenschaft

$$I(X_i, X_j | X_{i_1} \dots X_{i_k}) = I(X_j, X_i | X_{i_1} \dots X_{i_k}),$$

sowie $0 \leq I(X_i, X_j | X_{i_1} \dots X_{i_k})$. Sind X_i und X_j unabhängig von der Menge $\{X_{i_1}, \dots, X_{i_k}\}$ dann gilt $I(X_i, X_j | X_{i_1}, \dots, X_{i_k}) = 0$ und es folgt

$$\left(\frac{p_{X_i X_j X_{i_1} \dots X_{i_k}}}{p_{\cdot X_{i_1} \dots X_{i_k}}} \right) = \left(\frac{p_{X_i \cdot X_{i_1} \dots X_{i_k}}}{p_{\cdot X_{i_1} \dots X_{i_k}}} \right) \left(\frac{p_{\cdot X_j X_{i_1} \dots X_{i_k}}}{p_{\cdot X_{i_1} \dots X_{i_k}}} \right)$$

für alle $X_i, X_j, X_{i_1} \dots X_{i_k}$. Weiterhin ist $I(X_i, X_j | X_{i_1}, \dots, X_{i_k}) = 0$, wenn X_i unabhängig und X_j abhängig von $\{X_{i_1}, \dots, X_{i_k}\}$ ist und umgekehrt.

Sind X_i, X_j und $X_{i_1} \dots X_{i_k}$ kontinuierliche Zufallsvariablen mit der Wahrscheinlichkeitsverteilung $p(x_i, x_j, x_{i_1}, \dots, x_{i_k})$ und den Randdichten $p(x_i, \cdot, x_{i_1}, \dots, x_{i_k})$, $p(\cdot, x_j, x_{i_1}, \dots, x_{i_k})$ und $p(\cdot, \cdot, x_{i_1}, \dots, x_{i_k})$, dann ist die Partial Mutual Information

$$I(X_i, X_j | X_{i_1}, \dots, X_{i_k}) = \int p(x_i, x_j, x_{i_1}, \dots, x_{i_k}) \cdot \ln \frac{p(x_i, x_j, x_{i_1}, \dots, x_{i_k}) p(\cdot, \cdot, x_{i_1}, \dots, x_{i_k})}{p(x_i, \cdot, x_{i_1}, \dots, x_{i_k}) p(\cdot, x_j, x_{i_1}, \dots, x_{i_k})} dx_i dx_j dx_{i_1} \dots dx_{i_k}. \quad (7.4.2)$$

Die Gleichung (7.4.2) kann mit der Differential Entropie wie folgt formuliert werden

$$\begin{aligned} I(X_i, X_j | X_{i_1}, \dots, X_{i_k}) &= h(X_i, X_{i_1}, \dots, X_{i_k}) + h(X_j, X_{i_1}, \dots, X_{i_k}) \\ &\quad - h(X_{i_1}, \dots, X_{i_k}) - h(X_i, X_j, X_{i_1}, \dots, X_{i_k}). \end{aligned}$$

Die kontinuierliche Partial Mutual Information höherer Ordnung ist ebenso wie die gewöhnliche Partial Mutual Information invariant gegenüber streng monotoner Transformationen und damit robust gegenüber möglicher nichtlinearer Verformungen der Zeitreihen. Unter Vernachlässigung der Menge der Zufallsvariablen $\{X_{i_1}, \dots, X_{i_k}\}$ ergibt sich wieder die gewöhnliche Mutual Information:

$$I(X_i, X_j) = h(X_i) + h(X_j) - h(X_i, X_j).$$

7.4.1 Die Berechnung der Partial Mutual Information höherer Ordnung

Die Berechnung der Partial Mutual Information höherer Ordnung erfolgt über eine Verallgemeinerung der Methode von Kraskov, Stögbauer und Grassberger in [17] basierend auf den k nächsten Nachbar, wie im Abschnitt 7.3.1.2 beschrieben.

Ausgangspunkt für die Schätzung bezüglich der Ordnung k sind nun die Zeitreihen $\{x_i\}, \{x_j\}, \{x_{i_1}\}, \dots, \{x_{i_k}\}$, welche endliche Realisierungen von stationären ergodischen Prozessen $\{X_i\}, \{X_j\}, \{X_{i_1}\}, \dots, \{X_{i_k}\}$ sind. Gemäß der Methode für die erste Ordnung erfolgt für jeden Vektor $v_t = (x_i, x_j, x_{i_1}, \dots, x_{i_k})$, $t = 1, 2, \dots, T$, und festes k mit $1 \leq k < T$ die Bestimmung der Distanz $\varepsilon_k(t)$ zum k nächsten Nachbar. Demzufolge sind genau $k - 1$ Punkte mit der Distanz streng kleiner als $\varepsilon_k(t)$ in $\{v_{t^*}\}$ mit $t^* = 1, \dots, T$, $t^* \neq t$ und $N - k - 1$ Punkte streng größer als $\varepsilon_k(t)$. Die Distanzberechnung erfolgt mit der Maximumnorm. Nach Festlegung der Randvektoren $w_t = (x_i, x_{i_1}, \dots, x_{i_k})$, $t = 1, 2, \dots, T$ wird für jedes t die Anzahl $N_{x_i x_{i_1} \dots x_{i_k}}$ der Punkte in $\{w_{t^*}\}$ gezählt, welche eine Distanz streng kleiner als $\varepsilon_k(t)$ haben, wobei gilt $k \leq N_{x_i x_{i_1} \dots x_{i_k}} + 1$.

$$N_{x_i x_{i_1} \dots x_{i_k}}(t) = \# \{t^* \neq t : \|w_{t^*} - w_t\| < \varepsilon_k(t)\}$$

Für Schätzung von $I(X_i, X_j | X_{i_1}, \dots, X_{i_k})$ erfolgt die Bestimmung von $N_{x_j x_{i_1} \dots x_{i_k}}$ und $N_{x_{i_1} \dots x_{i_k}}$ für Randvektoren $(x_j, x_{i_1}, \dots, x_{i_k})$ und $(x_{i_1}, \dots, x_{i_k})$ wie folgt

$$N_{x_j x_{i_1} \dots x_{i_k}}(t) = \# \{t^* \neq t : \|w_{t^*} - w_t\| < \varepsilon_k(t)\}$$

$$N_{x_{i_1} \dots x_{i_k}}(t) = \# \{t^* \neq t : \|w_{t^*} - w_t\| < \varepsilon_k(t)\}$$

Somit ergibt sich für die Schätzung von $I(X_i, X_j | X_{i_1}, \dots, X_{i_k})$

$$\hat{I}(X_i, X_j | X_{i_1}, \dots, X_{i_k}) = \frac{1}{T} \sum_{t=1}^T \left(h_{N_{x_i x_{i_1} \dots x_{i_k}}(t)} + h_{N_{x_j x_{i_1} \dots x_{i_k}}(t)} - h_{N_{x_{i_1} \dots x_{i_k}}(t)} \right) - h_{k-1}.$$

Für $k = 1$ resultiert daraus die Partial Mutual Information erster Ordnung. Bei Vernachlässigung der Variablen x_{i_1}, \dots, x_{i_k} , das heißt $k = 0$, erhält man die gewöhnliche Mutual Information $I(X_i, X_j)$. In diesem Kontext folgt $N_{x_i x_{i_1} \dots x_{i_k}} = N_{x_i}$, $N_{x_j x_{i_1} \dots x_{i_k}}$, sowie $N_{x_{i_1} \dots x_{i_k}} = T - 1$.

$$\hat{I}(X_i, X_j) = \frac{1}{T} \sum_{t=1}^T (h_{N_{x_i}(t)} + h_{N_{x_j}(t)}) - h_{T-1} - h_{k-1}.$$

Die Partial Mutual Information höherer Ordnung kann auch mit der Methode basierend auf dem k -nächsten Nachbar geschätzt werden. Damit kann nun die Matrix $MI_{Exemplare}^{partial}$ angegeben werden. Diese Matrix wird auf der Basis der 7 Exemplare bestimmt, die aus der Clusterung mit dem Algorithmus Affinity Propagation resultierten. Die neue Matrix $MI_{Exemplare}^{partial}$ geht aus der Matrix $MI_{Exemplare}$ hervor, indem nur dann ein Matricelement von $MI_{Exemplare}$ geändert wird, wenn eine Pseudo Mutual Information vorliegt. Genauer, wenn eine Pseudo Mutual Information zwischen den Exemplaren i und j besteht, dann ersetzt man das entsprechende Matricelement $MI_{Exemplare}^{partial}(i, j) = \hat{I}(X_i, X_j | X_{i_1}, \dots, X_{i_k})$, andernfalls bleibt das Matricelement $MI_{Exemplare}(i, j)$ erhalten, dass heißt $MI_{Exemplare}^{partial}(i, j) = \hat{I}(X_i, X_j) = MI_{Exemplare}(i, j)$.

$$MI_{Exemplare}^{partial} = \begin{pmatrix} 0 & 0,1527 & 0,1476 & 0,1638 & 0,1313 & 0,1483 & 0,2716 \\ 0,1527 & 0 & 0,2548 & 0,1618 & 0,1833 & 0,3015 & 0,2098 \\ 0,1476 & 0,2548 & 0 & 0,1683 & 0,2193 & 0,1620 & 0,2039 \\ 0,1638 & 0,1648 & 0,1683 & 0 & 0,1876 & 0,1505 & 0,1574 \\ 0,1313 & 0,1833 & 0,2193 & 0,1876 & 0 & 0,1546 & 0,1607 \\ 0,1483 & 0,3015 & 0,1620 & 0,1505 & 0,1546 & 0 & 0,1468 \\ 0,2716 & 0,2098 & 0,2039 & 0,1574 & 0,1607 & 0,1468 & 0 \end{pmatrix} \quad (7.4.3)$$

Demnach entsteht die Matrix $MI_{Exemplare}^{partial}$ (7.4.3) auf analoger Weise zur Matrix $A_{Exemplare}^{partial}$ (6.3.5), nur das jetzt anstatt der Korrelation die Mutual Information verwendet wird.

Beim Vergleich der Matrizen $MI_{Exemplare}$ (7.3.14) und $MI_{Exemplare}^{partial}$ (7.4.3) ist zu erkennen, dass kaum ein Matricelement von $MI_{Exemplare}$ in $MI_{Exemplare}^{partial}$ erhalten geblieben ist, dass heißt, es sind viele Pseudo Mutual Informations aufgetreten. Dadurch enthält die Matrix $MI_{Exemplare}^{partial}$ nur noch relativ kleine Werte und damit eine kleine Varianz innerhalb der Werte.

8 Auswertung und Ergebnisse

In diesem Kapitel erfolgt eine Gegenüberstellung der Ergebnisse der partiellen Korrelationsmatrix und der Partial Mutual Information Matrix. Anschließend werden die Ergebnisse für diese Matrizen bezogen auf die Anwendung der Eigenvektor Zentralität dargelegt.

8.1 Vergleich der partiellen Korrelationsmatrix und der Partial Mutual Information Matrix

Mit der Verallgemeinerung der Partial Mutual Information können die partielle Korrelation k -ter Ordnung und die Partial Mutual Information der Ordnung k gegenüber gestellt werden.

Aus dem Vergleich der Matrizen $A_{Exemplare}$ (6.1.2), $A_{Exemplare}^{partial}$ (6.3.5), $MI_{Exemplare}$ (7.3.14) und $MI_{Exemplare}^{partial}$ (7.4.3) untereinander ist ersichtlich, dass die Werte in der Korrelationsmatrix $A_{Exemplare}$ überwiegend geringfügig kleiner sind als in der Mutual Information Matrix $MI_{Exemplare}$. Liegt ein hoher Korrelationskoeffizient zwischen zwei Exemplaren X_i und X_j vor, dann ergibt sich auch ein hoher Wert für die Mutual Information zwischen X_i und X_j . Grob formuliert hat die Korrelation und die Mutual Information eine ähnliche Charakteristik für den zugrundeliegenden fMRT-Datensatz.

Im vorigen Kapitel wurde festgestellt, dass bei der Verwendung der Mutual Information viele Pseudo Mutual Information aufgetreten sind. Im Gegensatz dazu wurden bei den Korrelationen weniger Scheinkorrelationen aufgedeckt. Damit enthält die Matrix $A_{Exemplare}^{partial}$ eine höhere Varianz der Werte als die Matrix $MI_{Exemplare}^{partial}$. Inhaltliche Schlussfolgerungen könnten somit vielleicht besser aus der Matrix $A_{Exemplare}^{partial}$ gezogen werden.

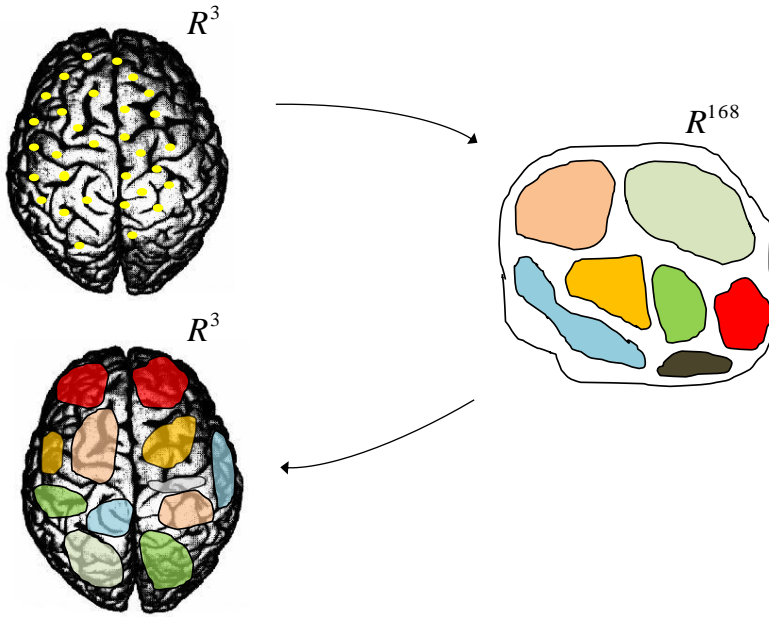


Abb. 8.2.1: Eine schematische Darstellung der Datenverarbeitung:

Zuerst erfolgt eine Vektorquantisierung des original Datensatzes, gekennzeichnet durch die gelb markierten Voxel im Gehirn, zu dem jeweils eine Zeitreihe existiert. Die daraus resultierenden Prototypen werden in Cluster zusammengefasst. Dabei bedeuten verschiedene Farben unterschiedliche Cluster. Anschließend können die gewonnenen Cluster auf den original Datenraum zurückgeführt und räumlich dargestellt werden, da die Voxel-Koordinaten zu jeder Zeitreihe bekannt sind.

8.2 Auswertung der Eigenvektor Zentralität

Vor der Berechnung der Eigenvektor Zentralität erfolgt eine Zurückführung der Exemplare auf die Prototypen und dann auf den originalen Datensatz, siehe schematische Darstellung in Abbildung 8.2.1. Damit werden aus den Matrizen $A_{Exemplare}^{partial}$ und $MI_{Exemplare}^{partial}$ zunächst die Matrizen $A_{Prototypen}^{partial}$ und $MI_{Prototypen}^{partial}$ generiert, welche nun die Größe 100×100 haben, da beim Neural Gas Algorithmus 100 Prototypen bestimmt wurden. Genauer entstehen die Matrizen $A_{Prototypen}^{partial}$ und $MI_{Prototypen}^{partial}$ wie folgt: Die Zurücktransformation erfolgt mit der Kenntnis, welcher Prototyp zu welchen Cluster und damit zu welchen Exemplar zugeordnet ist. Das bedeutet bezug nehmend auf die Korrelationsmatrix, jedes

Matrizelement in $A_{Prototypen}^{partial}$ erhält den entsprechenden Wert des Matrizelementes von $A_{Exemplare}^{partial}$. Gehören beispielsweise Prototyp p_n zum Exemplar e_k und Prototyp p_m zum Exemplar e_l , wobei $n, m \in \{1, \dots, 100\}$ und $k, l \in \{1, \dots, 7\}$, dann überträgt sich das Matrizelement $a_{k,l} \in A_{Exemplare}^{partial}$ auf das Matrizelement $a_{n,m} \in A_{Prototypen}^{partial}$.

Anschließend erfolgt in analoger Weise die Zurückführung der Prototypen auf die original fMRT-Zeitreihen. Damit resultieren die Matrizen $A_{Zeitreihen}^{partial}$ und $MI_{Zeitreihen}^{partial}$, welche sich aus den Matrizen $A_{Prototyp}^{partial}$ und $MI_{Prototyp}^{partial}$ ergeben. Anstatt der gewöhnlichen Korrelationsmatrix als Basis für die Berechnung der Eigenvektor Zentralität ist nun entweder $A_{Zeitreihen}^{partial}$ oder die Matrix $MI_{Zeitreihen}^{partial}$ die Ähnlichkeitsmatrix, welche nun eine Größe von 39460×39460 haben. Falls notwendig erfolgt eine Skalierung der Matrizen $A_{Zeitreihen}^{partial}$ und $MI_{Zeitreihen}^{partial}$, damit die Voraussetzung für die Berechnung der Eigenvektor Zentralität erfüllt ist.

Allerdings entstehen durch die Zurücktransformation in den Matrizen $A_{Zeitreihen}^{partial}$ und $MI_{Zeitreihen}^{partial}$ Blöcke mit gleichen Werten. Dadurch haben die Matrizen $A_{Zeitreihen}^{partial}$ und $MI_{Zeitreihen}^{partial}$ lediglich 21 verschiedene nicht diagonal Elemente, da die symmetrischen Matrizen $A_{Exemplare}^{partial}$ und $MI_{Exemplare}^{partial}$ nur diese 21 verschiedenen nicht diagonal Elemente haben. Folglich sind zu viele identische Matrizelemente in $A_{Zeitreihen}^{partial}$ und $MI_{Zeitreihen}^{partial}$ vorhanden und enthalten somit zu wenig Information. Die Berechnung der Eigenvektor Zentralität basierend auf den Matrizen $A_{Zeitreihen}^{partial}$ und $MI_{Zeitreihen}^{partial}$ bestätigt diese Aussage. Auch die Visualisierung des Zentralitätsmaßes im Gehirn zeigt, dass dadurch alle Voxel einen hohen Zentralitätswert erhalten. Daher ist diese Form der Zurückführung nicht sinnvoll.

Eine bessere Vorgehensweise ist die partiellen Korrelation zwischen jeweils zwei fMRT-Zeitreihen des Datensatzes unter der Bedingung der sieben Exemplare zu bestimmen. Im Detail ist damit jedes Matrizelement in der neuen Matrix $A^{partial}$ die partielle Korrelation $r_{X_i X_j | e_1 \dots e_7} \forall i, j \in \{1, \dots, n\}$, wobei $e_1 \dots e_7$ die Exemplare sind und n die Anzahl der Zeitreihen ist, hier $n = 39460$.

Die Bestimmung der Eigenvektor Zentralität für das vollständige Gehirn kann nun auf der neuen Matrix $A^{partial}$ erfolgen und mit der hauseigenen Software vom Max Planck Institut visualisiert werden. In den folgenden Abbildungen, die das Ergebnis der Eigenvektor Zentralität darstellen, zeigt der jeweilige Mittelpunkt des weißen Kreuzes die betrachtete Gehirnregion. Dabei sind die Gehirnregionen mit einer höheren Zentralität farbig gekennzeichnet. Zu erwarten wäre, dass höhere Zentralitäten im Kortex oder im

subkortikalen Kortex gefunden werden.

Die ersten zwei Abbildungen (8.2.2 und 8.2.3) zeigen die Auswertungen im Nullpunkt des Koordinatensystems, welcher mit CA (Commissure Anterior) bezeichnet wird. Anhand dieser zwei Bilder ist zu erkennen, dass die Verwendung der partiellen Korrelation und der marginalen Korrelation zu unterschiedlichen Ergebnissen in der Eigenvektor Zentralität führt. Auf der Grundlage der partiellen Korrelationsmatrix entstehen höhere Zentralitäten für andere Hirnregionen als bei der marginalen Korrelationsmatrix. Für eine genauere Analyse der Ergebnisse basierend auf der partiellen Korrelationsmatrix wird speziell der Hippocampus, der Orbitofrontale Kortex, der Retrosplenial Kortex und der Thalamus genauer betrachtet.

Im Hippocampus (siehe Abbildung 8.2.4) liegt eine höhere Zentralität besonderes in der linken Hemisphäre vor. Ebenso zeigt die Abbildung 8.2.5 höhere Zentralitäten im Orbitofrontalen Kortex und in Abbildung 8.2.6 die höhere Zentralität im Retrosplenialen Kortex. Für beide Matrizen sind höhere Zentralitätswerte im Thalamus (siehe Abbildung 8.2.7 und 8.2.8) zu finden. Zusammenfassend kann aus diesen Abbildungen folgende wichtige Erkenntnis hervorgehoben werden: Höhere Zentralitätswerte, welche mit der Eigenvektor Zentralität basierend auf den partiellen Korrelationen bestimmt wurden, sind im Kortex gefunden wurden. Dieses Resultat entspricht der oben genannten Erwartung und ist im Sinne der Neurowissenschaft durchaus plausibel.

Weiterhin stimmen diese Resultate mit einem unveröffentlichten Ergebnis bezüglich der Eigenvektor Zentralität von G. Lohmann überein. Dabei hat sie das Ergebnis der Eigenvektor Zentralität mit dem zweiten Eigenvektor ausgewertet. Es zeigten sich Übereinstimmungen höherer Zentralitäten gleicher Hirnregionen, obwohl verschiedene Methoden zur Bestimmung der Zentralitätswerte gewählt wurden.

Zukünftig könnte analog auch die Auswertung der Partial Mutual Information für das vollständige Gehirn erfolgen.

Ein gewisser Nachteil dieser Methode bezogen auf die Korrelation ist, dass man für jedes Matrixelement von $A^{partial}$ die partielle Korrelation ermittelt und damit die Berechnungszeit enorm ist. Ein zukünftiger Ausweg könnte möglicherweise sein, dass man nur spezielle Elemente in der Korrelationsmatrix mit der partiellen Korrelation $r_{X_i X_j | e_1 \dots e_7}$ $\forall i, j \in \{1, \dots, n\}$, wobei $e_1 \dots e_7$ wieder die Exemplare sind und n die Anzahl der Zeitreihen, berechnet und die restlichen Matrixelemente mit der gewöhnlichen Korrelation $r_{X_i X_j}$

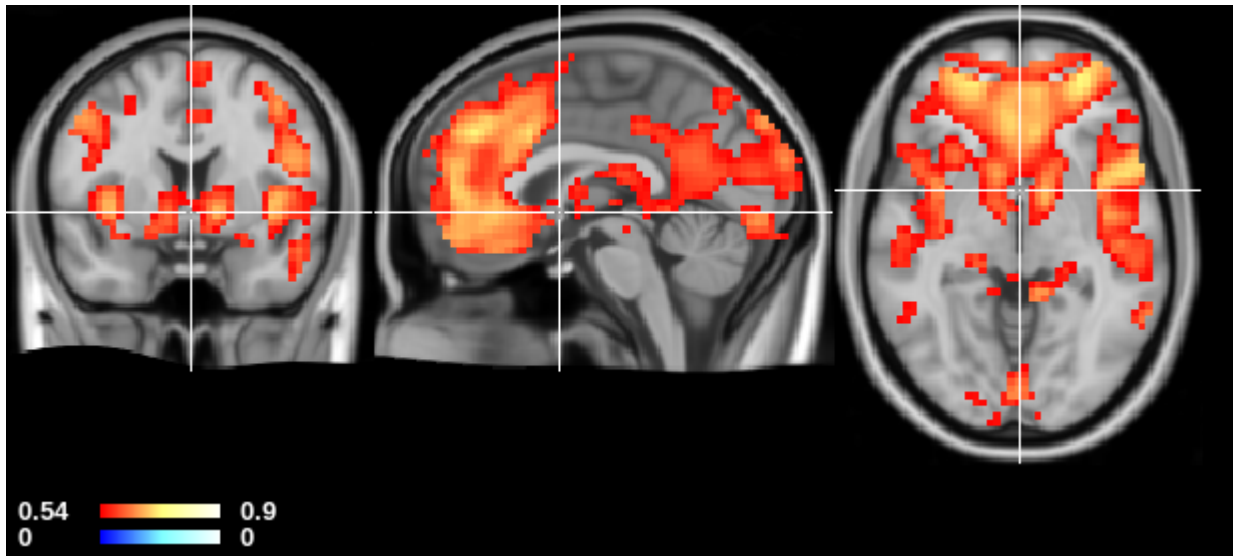


Abb. 8.2.2: Darstellung der Ergebnisse basierend auf der partiellen Korrelationsmatrix im Nullpunkt des Koordinatensystems, das heißt im Punkt Anterior Commissure. Der Begriff Anterior Commissure bezeichnet ein Nervenfaserbündel, welches rechte und linke Hemisphäre verbindet.

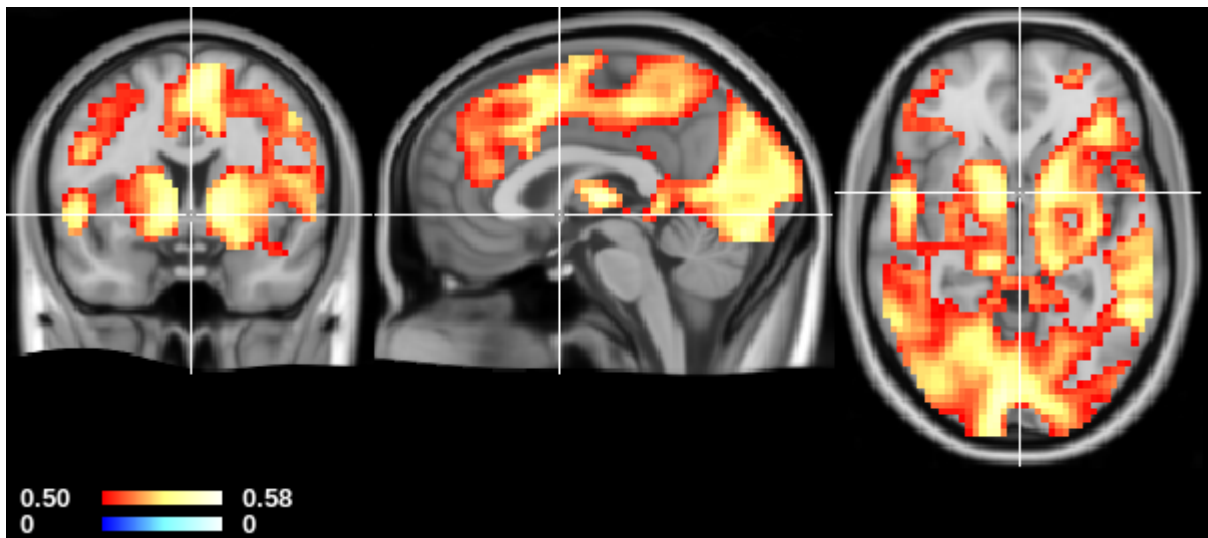


Abb. 8.2.3: Darstellung der Ergebnisse basierend auf der gewöhnlichen Korrelationsmatrix im Nullpunkt des Koordinatensystems.

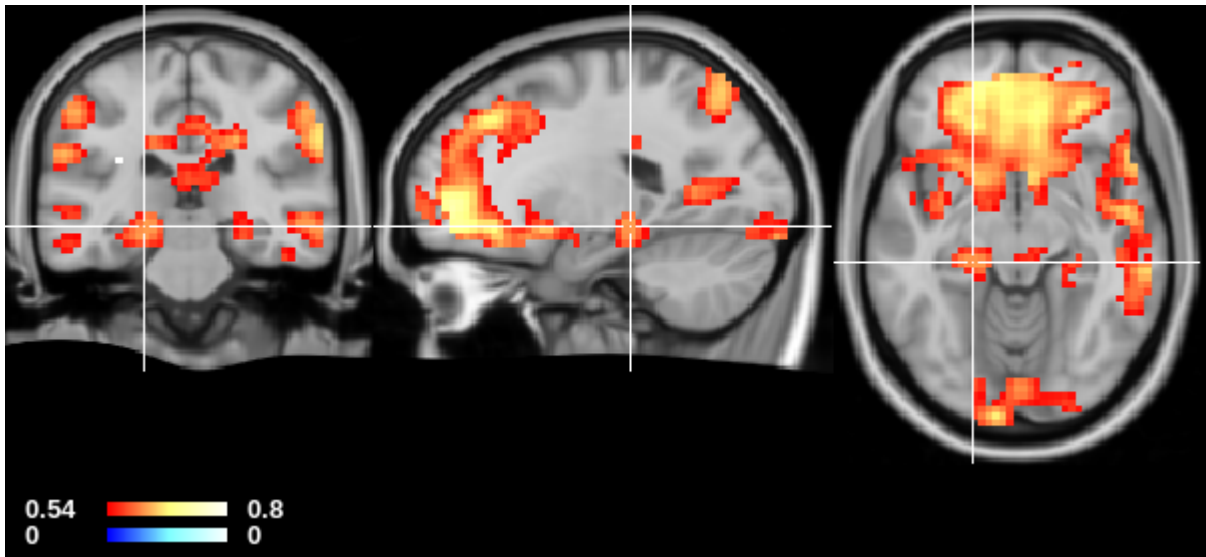


Abb. 8.2.4: Höhere Zentralitätswerte auf der Grundlage der partiellen Korrelationen im Hippocampus, welcher zuständig ist für die Gedächtniskonsolidierung.

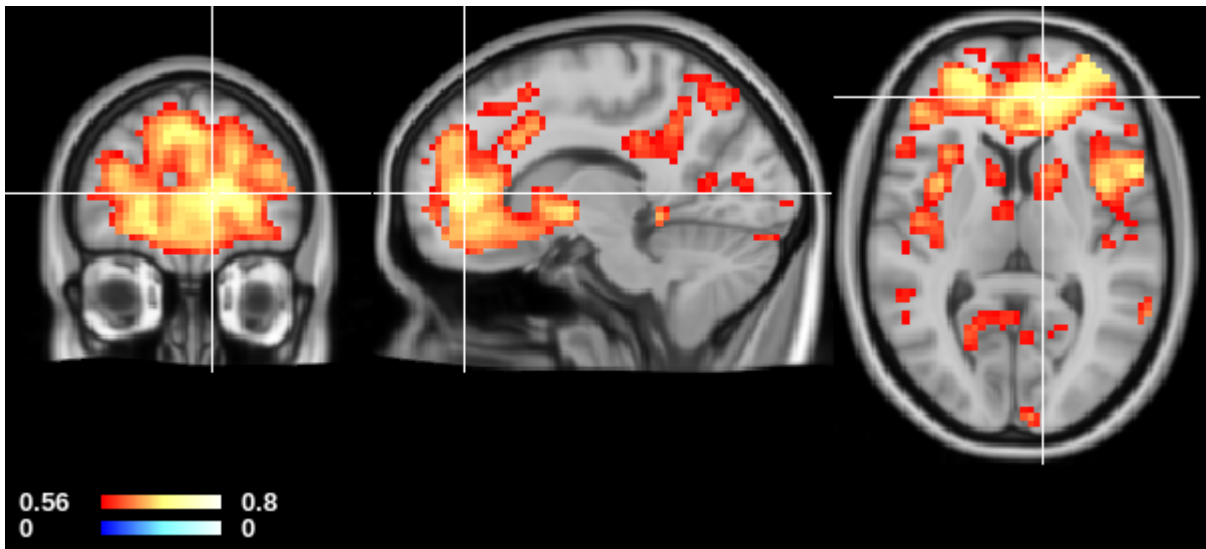


Abb. 8.2.5: Darstellung der höheren Zentralitätswerte auf der Grundlage der partiellen Korrelationsmatrix im Orbitofrontaler Kortex, welcher für die Emotions- und Impulskontrolle zuständig ist.

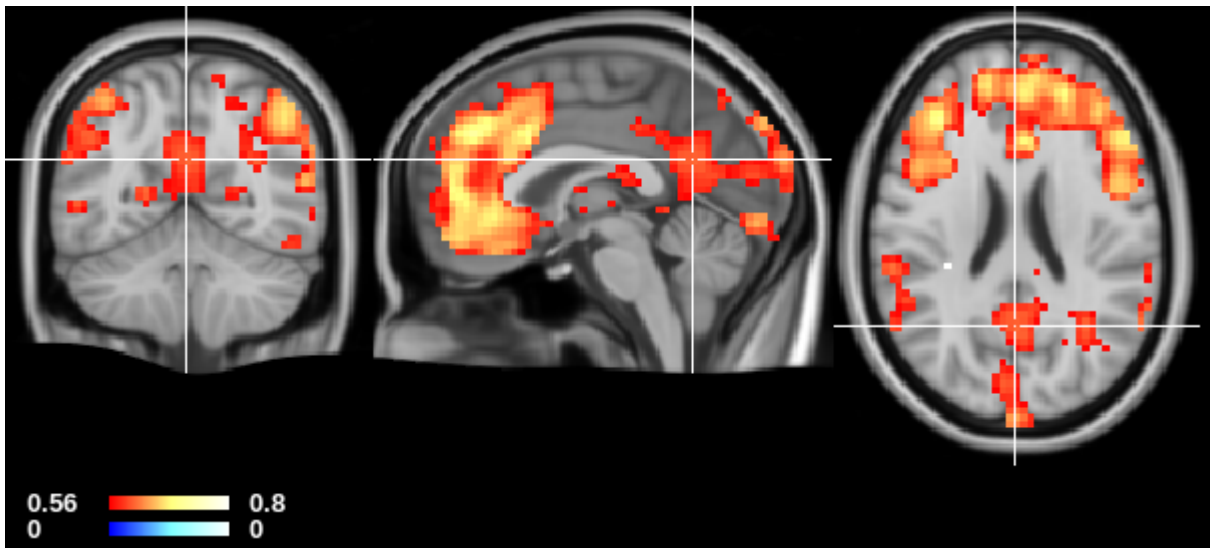


Abb. 8.2.6: Darstellung höherer Zentralitäten basierend auf der partiellen Korrelationsmatrix im Retrosplenial Kortex.

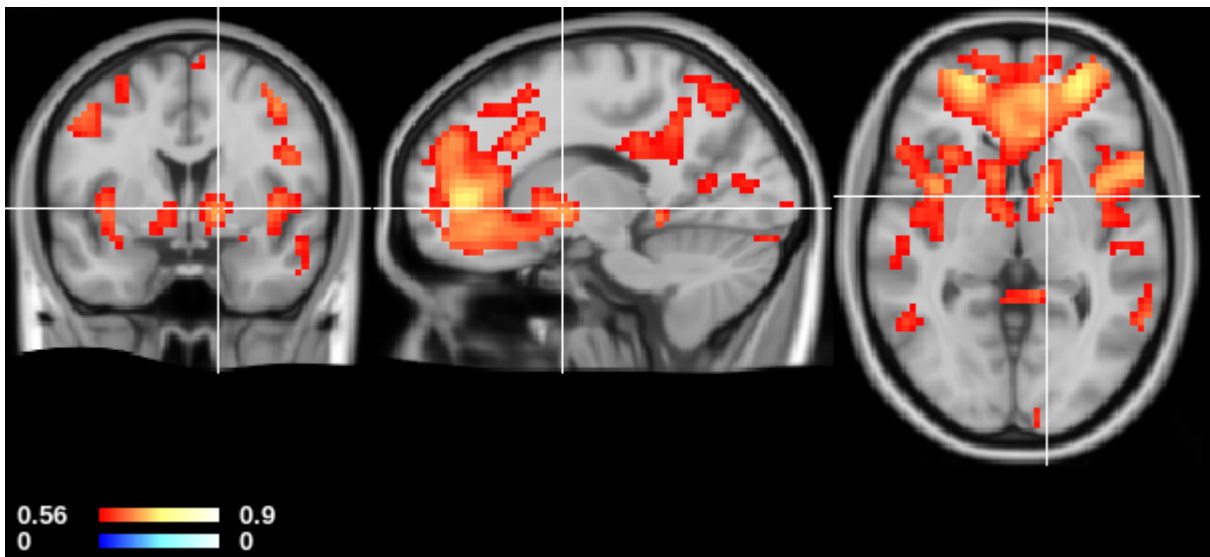


Abb. 8.2.7: Hohe Zentralitätswerte im Bereich des Thalamus, wobei hier die Ähnlichkeitsmatrix für die Eigenvektor Zentralität die partielle Korrelationsmatrix ist.

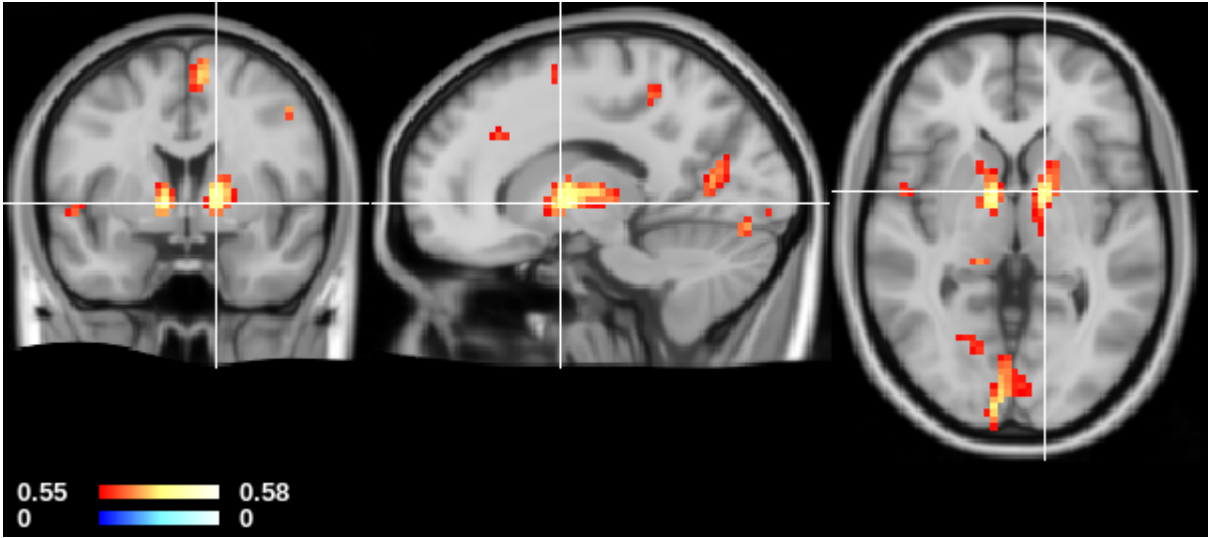


Abb. 8.2.8: Hohe Zentralitätswerte im Bereich des Thalamus, wobei hier die Ähnlichkeitsmatrix für die Eigenvektor Zentralität die gewöhnliche Korrelationsmatrix ist.

$\forall i, j \in \{1, \dots, n\}$ bestimmt. Unter Ausnutzung des Scheinkorrelationsmaßes und der Matrix $A_{Exemplare}^{partial}$ kann genau angegeben werden, zwischen welchen fMRT-Zeitreihen die partielle Korrelation bestimmt wird, da die Zuordnung zu den Prototypen und damit zu fMRT-Zeitreihen bekannt ist. Das bedeutet, in der so neu resultierenden Matrix $A^{partial}$ wird zwischen zwei fMRT-Zeitreihen X_i und X_j nur dann die partielle Korrelation $r_{X_i X_j | e_1 \dots e_7}$ bestimmt, wenn zwischen den entsprechenden zwei Exemplaren eine Scheinkorrelation vorliegt. Da speziell in der Matrix $A_{Exemplare}^{partial}$ nur fünf Scheinkorrelationen aufgedeckt wurden, ist damit nur zwischen den fMRT-Zeitreihen die partielle Korrelation zu bestimmen, die zu den Exemplaren zugeordnet sind.

9 Zusammenfassung und Diskussion

Im Vordergrund der Arbeit stand die Untersuchung der partiellen Korrelation und der Partial Mutual Information zur Untersuchung von Netzwerkstrukturen im Gehirn. Damit sollen indirekte Verbindungen im vollständigen Gehirn aufgedeckt werden. Diese Untersuchungen beziehen sich ausschließlich auf den gleichen Datensatz. Weiter sollte die partielle Korrelation und die Partial Mutual Information auf die Eigenvektor Zentralität angewendet werden.

Um die partielle Korrelation und die Partial Mutual Information auf dem gesamten Datensatz zu analysieren, wurden die fMRT-Zeitreihen mit einer Vektorquantisierung komprimiert. Hierfür nutzte man den Neural Gas Algorithmus, welcher mit 100 Prototypen die fMRT-Zeitreihen (des vollständigen Datensatz) repräsentieren soll. Die spezielle Prototypenanzahl wählte man aufgrund der Kenntnis der ungefähren Anzahl der Gehirnregionen. Allerdings entsprechen diese Prototypenvektoren Mittelungen der Datenvektoren, womit ein Glättungseffekt in den Prototypen entsteht. Die Glättung der hochfrequenten Signalanteile stellt jedoch einen Informationsverlust dar. Deshalb wurden anstatt der Prototypen, resultierend aus dem Neural Gas Algorithmus, originale fMRT-Zeitreihen für das weitere Vorgehen verwendet, welche am besten auf die Prototypen abgebildet werden. Außer der Datenkompression hat die Vektorquantisierung den Vorteil, dass die resultierenden Prototypen einen speziellen ε Radius haben. Damit löst es das Problem, dass die partiellen Korrelationen nicht zwischen den direkt benachbarten Zeitreihen bestimmt werden können, aufgrund der Ähnlichkeit der direkt benachbarten fMRT-Zeitreihen. Der Neural Gas Algorithmus ist für die fMRT-Zeitreihen sehr geeignet, da er in kurzer Zeit gute Repräsentanten für den fMRT-Datensatz liefert.

Zusätzlich bestand die Annahme, dass sich ähnliche Prototypen in Cluster vereinigen lassen, was eine weitere erwünschte Datenreduktion zur Folge hätte. Der hier verwendete Cluster-Algorithmus Affinity Propagation wird mit einem Präferenzwert und einer Ähnlichkeitsmatrix der Prototypen initialisiert. Der Vorteil von Affinity Propagation liegt

darin, dass dieser Algorithmus die Möglichkeit bietet, stabile Konfigurationen (Clusteranzahlen) zu bestimmen. Damit erhält der Anwender die Information, wie der Präferenzwert für eine stabile Clusterlösung zu wählen ist. Die Anwendung des Cluster-Algorithmus Affinity Propagation ergab bei einem Präferenzwert mit maximaler Clusteranzahl eine stabile Clusterlösung mit genau sieben Clusterzentren (Exemplaren). Unter Ausnutzung der Voxelkoordinaten und der Zurückverfolgung auf das Gehirn zeigten sich gut separierte und kompakte Clusterverteilungen. Zudem hat Affinity Propagation den weiteren Vorteil, dass die Berechnungszeit auf der Basis der Prototypen nur wenige Sekunden beträgt. Von zukünftigem Interesse wäre die Beurteilung der Clusterverteilungen mit speziellen Validierungsmaßen, sowie die Verwendung verschiedener Ähnlichkeitsmatrizen zur Initialisierung des Algorithmus. Dadurch könnten sich Clusterlösungen ergeben, die auch nach den hochfrequenten Signalanteilen geclustert sind. Weiter sollten die Clusterlösungen mit der hauseigenen Software LIPSIA vom Max Planck Institut visualisiert werden, um die Ergebnisse besser beurteilen zu können.

Basierend auf den sieben Exemplaren wurde die Korrelationsmatrix nach Pearson, Spearman und Kendall bestimmt. Ein Vergleich dieser Matrizen zeigte nur geringfügige Unterschiede, weshalb die Schätzung des Korrelationskoeffizienten weiterhin nach Pearson erfolgen kann.

Mit der partiellen Korrelation höherer Ordnung wurden die Scheinkorrelationen zur Aufdeckung indirekter Verbindungen im Gehirn genauer untersucht. Dabei wurde ein Scheinkorrelationsmaß entwickelt, welches aussagt, ab wann der Unterschied zwischen der marginalen Korrelation und partiellen Korrelation relevant ist und damit eine indirekte Verbindung zwischen zwei fMRT-Zeitreihen vorliegt. Mit dieser Kenngröße entstand nun die neue partielle Korrelationsmatrix, welche aus der Korrelationsmatrix basierend auf den sieben Exemplaren hervorging. Diese neue Korrelationsmatrix hat damit die Größe 7×7 und enthält abgesehen von der Hauptdiagonale nur 21 verschiedene Matrixelemente. Nach einer Zurückführung der Exemplare auf die Prototypen und dann auf die fMRT-Zeitreihen hätte damit die neue Korrelationsmatrix, der Größe 39460×39460 , lediglich diese 21 verschiedenen Matrixelemente. Diese so konstruierte neue Korrelationsmatrix hat damit zu viele identische Werte und ist damit ungeeignet für die Analyse von Netzwerkstrukturen im Gehirn und damit auch für die Anwendung der Eigenvektor Zentralität. Daher sollte hier entweder eine andere Form der Zurückführung oder die Bestimmung der partiellen Korrelationsmatrix auf anderer Weise erfolgen.

Daraufhin wurden für das vollständige Gehirn die partiellen Korrelationen zwischen je zwei fMRT-Zeitreihen unter der Bedingung der sieben Exemplare ermittelt. Die Eigenvektor Zentralität basierend auf dieser neuen partiellen Korrelationsmatrix zeigte, dass nur die gewünschten Kortex-Areale hohe Zentralitätswerte erhielten. Beim Vergleich der Zentralitätswerte resultierend auf der Grundlage der marginalen Korrelationsmatrix und der partiellen Korrelationsmatrix wurde festgestellt, dass die Zentralitäten bezüglich der partiellen Korrelation im Sinne der Neurowissenschaft die interessanteren Ergebnisse liefert. Weiter hat sich herausgestellt, dass dieses Resultat mit einem neuen unveröffentlichten Ergebnis der Eigenvektor Zentralität von Frau Lohmann übereinstimmt. Daher scheint diese Art zur Konstruktion der partiellen Korrelationsmatrix und damit auch die vorigen Schritte, welche zur Entstehung der Exemplare führten, durchaus geeignet zu sein. Ein weiterer Vorteil dieser Vorgehensweise ist, dass die partiellen Korrelationen nicht nur auf spezielle Gehirnregionen, sondern auf dem vollständigen Gehirn bestimmt werden können.

Analog zu den Korrelationen wurde in dieser Arbeit auch die Mutual Information auf die fMRT-Zeitreihen angewendet. Somit wurden nicht nur die linearen Abhängigkeiten zwischen den fMRT-Zeitreihen betrachtet, sondern auch die nicht linearen Abhängigkeiten, da die Form der Abhängigkeit zwischen den fMRT-Zeitreihen unbekannt ist. Die Bestimmung der Mutual Information erfolgte über eine Schätzung basierend auf den k -nächsten Nachbar. Diese Methode hat allerdings den Nachteil, dass kein t-Test als Signifikanztest angewendet werden kann, da diese resultierenden Werte nicht in Fisher'sche Z-Werte transformiert werden können. Somit wurde ein neues Verfahren speziell für die Mutual Information als Ersatz für den Signifikanztest geschaffen.

Auf der Grundlage der sieben Exemplare konnte nun eine analoge Matrix zur Korrelationsmatrix angegeben werden. Die Gegenüberstellung dieser zwei Matrizen zeigte, dass einzelne Matrixelemente trotz unterschiedlicher Verfahren den gleichen Wert annehmen. Hingegen haben die anderen Matrixelemente nur geringe Unterschiede. Daraus kann allerdings nicht gefolgert werden, dass die Abhängigkeit zwischen den fMRT-Zeitreihen von linearer Natur ist. Sondern, dass die Mutual Information nicht ausreicht um die Nichtlinearitäten aufzudecken. Ein kleiner Nachteil gegenüber der Korrelation ist, dass die Berechnungszeit für die Mutual Information etwas höher liegt. Die Ursache dafür ist, dass die Schätzung der Mutual Information eine größere Anzahl von Berechnungsschritten benötigt als die Schätzung der Korrelation.

Weiterhin wurde analog zur Scheinkorrelation der Begriff Pseudo Mutual Information eingeführt. Damit war es auch notwendig ein Maß analog zum Scheinkorrelationsmaß zu entwickeln. Um nun die partielle Korrelation höherer Ordnung der Partial Mutual Information höherer Ordnung gegenüber stellen zu können, wurde die Partial Mutual Information erster Ordnung verallgemeinert dargelegt. Dadurch konnte man die Partial Mutual Information Matrix basierend auf den sieben Exemplaren bestimmen. In der so resultierenden Partial Mutual Information Matrix sind wesentlich kleinere Werte enthalten als in der Mutual Information Matrix und damit ist auch die Varianz innerhalb dieser Matrixelemente deutlich geringer. Für inhaltliche Schlussfolgerungen wäre eventuell eher die Mutual Information geeignet. Das oben erwähnte Problem bei der Zurückführung bezüglich der Korrelationsmatrix tritt auch bei der Mutual Information auf. Zukünftig sollte die Konstruktion der Partial Mutual Information Matrix analog zur partiellen Korrelationsmatrix erfolgen oder die Idee der Zurücktransformation, wie im Abschnitt 8.2 erläutert, getestet werden. Damit könnte es möglich sein, die partiellen Korrelationen und Partial Mutual Information auf dem gesamten Datensatz effektiver berechnen zu können.

A Notationen

X	Menge
V_i	Voronoi Zellen
$w_i(v)$	Referenzvektor
d	Ähnlichkeitsmaß
$P(v)$	Dichteverteilung
$k_i(v, w)$	Nachbarschaftsrang
ϵ	Schrittweite
g_λ	Gaußfunktion
$c(\lambda)$	Normalisierungsfaktor
E_{ng}	Kostenfunktion
1_{WTA}	Winner-Takes-all Funktion
$\varrho(w_i)$	Dichte der Prototypen
dim	intrinsische Dimension
Σ	Kovarianzmatrix
r_{XY}	Korrelations zwischen X und Y nach Pearson
r_{XY}^s	Korrelations zwischen X und Y nach Spearman
r_{XY}^k	Korrelations zwischen X und Y nach Kendall
$r_{XY Z=z}$	bedingte Korrelation
$r_{XY Z}$	partielle Korrelation
$f_{XY Z}$	bedingte Verteilung
$N(\mu, \Sigma)$	Normalverteilung
W	Wishart-Matrix
$W_k(\Sigma, n)$	Wishart Verteilung

$W_k(\Sigma, n, \lambda)$	Wishart Verteilung
\bar{x}	Mittelwert
\tilde{A}	skalierte Ähnlichkeitsmatrix mit der Korrelation nach Pearson
$A_{Exemplare}$	Korrelationsmatrix nach Pearson basierend auf den Exemplaren
$A_{Exemplare}^s$	Korrelationsmatrix nach Spearman basierend auf den Exemplaren
$A_{Exemplare}^k$	Korrelationsmatrix nach Kendall basierend auf den Exemplaren
$A_{Exemplare}^{partial}$	partielle Korrelationsmatrix nach Pearson basierend auf den Exemplaren
$A_{Prototyp}^{partial}$	partielle Korrelationsmatrix nach Pearson basierend auf den Prototypen
$A_{Zeitreihen}^{partial}$	partielle Korrelationsmatrix nach Pearson basierend auf den Zeitreihen
$MI_{Exemplare}$	Mutual Information Matrix basierend auf den Exemplaren
$MI_{Exemplare}^{partial}$	Partial Mutual Information Matrix basierend auf den Exemplaren
$MI_{Prototyp}^{partial}$	Partial Mutual Information Matrix basierend auf den Prototypen
$MI_{Zeitreihen}^{partial}$	Partial Mutual Information Matrix basierend auf den Zeitreihen
S_n	Schätzung der Kovarianzmatrix
H_n	Entropiefunktion
$H(X)$	Entropie von X
$H(X A)$	bedingte Entropie von X unter dem Ereignis A
Ω	Ereignisraum
$I(E)$	Information eines Ereignisses E
$I(X, Y)$	Mutual Information
$\hat{I}(X, Y)$	Schätzung der Mutual Information
$I(X, Y Z)$	Partial Mutual Information
$\hat{I}(X, Y Z)$	Schätzung der Partial Mutual Information
$h(X)$	Differential Entropie
B	Betafunktion
ψ	Digammafunktion
h_n	harmonische Zahl
ε_k	Distanz zum k nächsten Nachbar

Abbildungsverzeichnis

1.0.1	Diese Abbildung stellt eine indirekte schematisch Verbindung dar: Ein hoher Korrelationskoeffizient $r_{x_1x_2}$ zwischen den Voxeln x_1 und x_2 kann aufgrund des Einflusses von Voxel x_3 entstehen.	3
2.1.1	BOLD-Zeitreihen von einer aktivierten bzw. nicht aktivierten Hirnregion: Die Zeitreihe im Voxel A entspricht einer aktivierten Hirnregion und die Zeitreihe im Voxel B einer nicht aktivierten Hirnregion.	6
3.2.1	Die “region of interest”: Die Maske beinhaltet ≈ 40000 Voxels. Die Talairach Koordinaten für die Schnittposition sind $(0, 0, 0)$. [19]	10
3.2.2	Paarweise t-Test zwischen sattten und hungrigen Probanden mit linearer Korrelation für die Eigenvektor Zentralität: Ergebnis mit $p < 0.05$. Die Zentralitätswerte der Hirnregion Precuneus sind signifikant höher im hungrigen Zustand. Andere Regionen zeigen keine signifikanten Effekte. Die Talairach-Koordinaten in der Schnittposition sind $(-2, -50, 56)$. [19]	12
4.3.1	Darstellung des Glättungseffektes entstehend durch den Neural Gas Algo- rithmus: Die blaue Zeitreihe ist der Prototyp, welcher vom Neural Gas Algorithmus ermittelt wurde und die rote ist die für diesen Prototypen am besten approximierte Zeitreihe.	17
5.2.1	Präferenzwerte für stabile Phasen Dieser Wert kann auf der abszissen Achse abgelesen werden und die ordinaten Achse gibt die Clusteranzahl an.	21

5.2.2	Geclusterten fMRT-Prototypen mit den Exemplaren. Jede Farbe stellt ein Cluster dar und die schwarzen Zeitreihen die jeweiligen Clusterzentrum.	21
5.2.3	Clusterverteilung im vollständigen Gehirn, gemäß der Clusterlösung und der Rückabbildung bezüglich der Voxelkoordinaten.	22
5.2.4	Clusterbildung in der XY-Ebene.	22
5.2.5	Clusterbildung in der XZ-Ebene.	23
5.2.6	Clusterbildung in der YZ-Ebene	23
7.2.1	Grafische Darstellung der Mutual Information für die Zufallsvariablen X und Y	47
7.3.1	Grafische Darstellung der Partial Mutual Information für die Zufallsvariablen X und Y unter der Bedingung Z	48
7.3.2	grafische Darstellung von N_{xz}	54
7.4.1	Grafische Darstellung der Partial Mutual Information 2-ter Ordnung	59
8.2.1	Eine schematische Darstellung der Datenverarbeitung: Zuerst erfolgt eine Vektorquantisierung des original Datensatzes, gekennzeichnet durch die gelb markierten Voxel im Gehirn, zu dem jeweils eine Zeitreihe existiert. Die daraus resultierenden Prototypen werden in Cluster zusammengefasst. Dabei bedeuten verschiedene Farben unterschiedliche Cluster. Anschließend können die gewonnenen Cluster auf den original Datenraum zurückgeführt und räumlich dargestellt werden, da die Voxel-Koordinaten zu jeder Zeitreihe bekannt sind.	64
8.2.2	Darstellung der Ergebnisse basierend auf der partiellen Korrelationsmatrix im Nullpunkt des Koordinatensystems, dass heißt im Punkt Anterior Commissure. Der Begriff Anterior Commissure bezeichnet ein Nervenfaserbündel, welches rechte und linke Hemisphäre verbindet.	67
8.2.3	Darstellung der Ergebnisse basierend auf der gewöhnlichen Korrelationsmatrix im Nullpunkt des Koordinatensystems.	67
8.2.4	Höhere Zentralitätswerte auf der Grundlage der partiellen Korrelationen im Hippocampus, welcher zuständig ist für die Gedächtniskonsolidierung.	68

8.2.5 Darstellung der höheren Zentralitätswerte auf der Grundlage der partiellen Korrelationsmatrix im Orbitofrontaler Kortex, welcher für die Emotions- und Impulskontrolle zuständig ist.	68
8.2.6 Darstellung höherer Zentralitäten basierend auf der partiellen Korrelationsmatrix im Retrosplenial Kortex.	69
8.2.7 Hohe Zentralitätswerte im Bereich des Thalamus, wobei hier die Ähnlichkeitsmatrix für die Eigenvektor Zentralität die partielle Korrelationsmatrix ist.	69
8.2.8 Hohe Zentralitätswerte im Bereich des Thalamus, wobei hier die Ähnlichkeitsmatrix für die Eigenvektor Zentralität die gewöhnliche Korrelationsmatrix ist.	70

Literaturverzeichnis

- [1] *Persönliche Kommunikation PD. Dr. Lohmann.*
- [2] *Netzwerkanalyse*, tcs.rwth-aachen.de/lehre/Netzwerkanalyse/SS2004/zentralaus.pdf, 19.12.2011. rwth aachen, 2004.
- [3] *Zusammenhangsstrukturen*, www.stat.uni-muenchen.de/leiten/Lehre/Material/Mult_07/ZusamStrukt.pdf, 19.12.2011. uni-münchen, 2007.
- [4] Michael Biehl, Barbara Hammer, Michel Verleysen, and Thomas Villmann. *Similarity-Based Clustering, Recent Developments and Biomedical Applications*. Springer, 2009.
- [5] Bonacich and Lloyd. Calculating status with negative relations. *Social Networks*, 23:331–338, 2004.
- [6] Jürgen Bortz and Christof Schuster. *Statistik für Human- und Sozialwissenschaftler*. Springer, 2010.
- [7] Rainer M. Bösel. *Das Gehirn, Ein Lehrbuch der funktionellen Anatomie für die Psychologie*. Kohlhammer, 2006.
- [8] Johannes Dreesman. *Zur statistischen Inferenz in Markov-Feldern, Markov-Chain-Monte-Carlo-Verfahren und Modelle mit räumlich variierender Koeffizienten*. Herbert Utz Verlag, 1998.
- [9] D. Francios, F. Rossi, V. Wertz, and M. Verleysen. Resampling methods for parameter-free and robust feature selection with mutual information. *Neurocomputing* 70, pages 1276–1288, 2006.
- [10] Stefan Frenzel and Bernd Pompe. Partial mutual information for coupling analysis of multivariate time series. *Physical Review Letters*, 99:204101–1 –204101–4, 2007.
- [11] Brendan J. Frey and Delbert Dueck. Clustering by passing messages between data points. Technical report, American Association for the Advancement of Science, 2007.

- [12] Peter Grassberger. Entropy estimates from insufficient samplings. Technical report, John-von-Neumann Institute for Computing, 2008.
- [13] W. Hackbusch. *Iterative Lösung grosser schwachbesetzter Gleichungssysteme*. Teubner, 1991.
- [14] Hartung and Elpelt. *Multivariate Statistik, Lehr- und Handbuch der angewandten Statistik*. Oldenburg, 1992.
- [15] Harro Heuser. *Lehrbuch der Analysis Teil 1*. Teubner-Verlag, 1998.
- [16] D. Jungnickel. *Codierungstheorie*. Spektrum Verlag, 1995.
- [17] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. Technical report, John-Neumann Institut for Computing, Forschungszentrum Jülich, 2008.
- [18] Mandy Lange. *Diplomarbeit, Untersuchung frequenzselektiver Filter für fMRT-Daten*. Hochschule Mittweida, 2008.
- [19] Gabriele Lohmann, Daniel S. Margulies, Annette Horstmann, Burkhard Pleger, and et al. Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain. *PLOSone*, 2010.
- [20] Stephen M., Smith, Karla Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, and et al. Network modelling methods for fMRI. *NeuroImage* 54, pages 875–891, 2011.
- [21] Guillaume Marrelec, Alexandre Krainik, Hugues Duffau, Melanie Pelegriani-Issac, and et al. Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage* 32, pages 228–237, 2006.
- [22] Thomas M. Martinetz, Stanislav G. Berkovich, and Klaus J. Schulten. "neural-gas" network for vector quantization and its application to time-series prediction. *IEEE Transactions on neural network*, 4:558–569, 1993.
- [23] Christoph Memmel. *Schätzrisiken in der Portfoliotheorie: Auswirkungen und Möglichkeiten der Reduktion*. EUL Verlag, 2004.
- [24] Karsten Müller. *Die Anwendung von Spektral- und Wavletanalyse zur Untersuchung der Dynamik von BOLD-Zeitreihen verschiedener Hirnareale*. MPI SERIES, 2006.
- [25] Bernd Pompe. *Verfahren der stochastischen Signalverarbeitung*. Institut für Physik der Universität Greifswald, 2009.

- [26] Dr. V. Pieper Dr. J. Tiedge Prof. Dr. O. Beyer, H. Hackel. *Wahrscheinlichkeitsrechnung und mathematische Statistik*. B. G. Teubner Verlagsgesellschaft K.-G., 1991.
- [27] Nalini Ravishanker and Dipak K. Dey. *A first course in linear model theory*. CHAPMAN&HALL/CRC, 2002.
- [28] Scheck. *Theoretische Physik 1, Mechanik, Von den Newton'schen Gesetzen zum deterministischen Chaos*. Springer Lehrbuch, 2003.
- [29] H. Schneider-Obermann. *Kanalkodierung*. Vieweg, 1998.
- [30] Werner A. Stahel. *Statistische Datenanalyse, Eine Einführung für Naturwissenschaftler*. Vieweg, 2008.
- [31] Prof. Dr. Tutz. *Multivariate Verfahren*. Ludwig-Maximilians- Universität München, Institut für Statistik,, 2001.
- [32] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, Bernard Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15:273 – 289, 1002.
- [33] T. Villmann, T. Geweniger, M. Kästner, and M. Lange. Theory of fuzzy neural gas for unsupervised vector quantization. *eingereicht bei der Konferenz ICAISC (International Conference on Artificial Intelligence and Soft Computing)*, 2011.
- [34] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of ‘small-world’ networks. *letters to nature*, 393:440–442, 1998.
- [35] D. Welsh. *Codes und Cryptography*. Oxford Science Publication, 1998.